



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 116879**

**TO: James Schultz**  
**Location: REM-2D18/2C18**  
**Art Unit: 1635**  
**Tuesday, March 23, 2004**  
**Case Serial Number: 10/054313**

**From: Paul Schulwitz**  
**Location: Biotech-Chem Library**  
**REM-1A65**  
**Phone: (571)272-2527**

**[paul.schulwitz@uspto.gov](mailto:paul.schulwitz@uspto.gov)**

### **Search Notes**

Examiner Schultz,

See attached results.

If you have any questions about this search feel free to contact me at any time.

Thank you for using STIC search services!

Paul Schulwitz  
Technical Information Specialist  
STIC Biotech/Chem Library  
(571)272-2527

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 22, 2004, 21:38:16 ; Search time 2831 Seconds  
(without alignments)  
4378.703 Million cell updates/sec

Title: US-10-054-313-1  
Perfect score: 1546  
Sequence: 1 MSWLLFLAHRVALAALPCRR.....FIGNEADRLAREGAKQSD 286

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -MODEL=frame\_p2n.model -DEV=xlh  
-Q/cgn2\_1/USFT0.spool/US10054313/runat\_22032004\_140415\_28070/app\_query.fasta\_1.455  
-DB=GenEmbl -QFT0=fastap -SUFFIX=sp2n.rge -MINWATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10054313@cgn1\_1\_2372 @runat\_22032004\_140415\_28070 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*

29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.pod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1526	98.7	1158	9	BC002973 Homo sapi
2	1526	98.7	1168	9	AF039652 Homo sapi
3	1526	98.7	1604	9	AK075490 Homo sapi
4	1523	98.5	1150	6	BD223780 Human RNA
5	1522	98.4	1113	9	AF048994 Homo sapi
6	1522	98.4	1147	9	AF048995 Homo sapi
7	1518	98.2	861	9	AJ224117 Homo sapi
8	1457.5	94.3	2120	9	AK096913 Homo sapi
9	1197.5	77.5	1430	10	AF048993 Mus muscu
10	1193.5	77.2	1409	10	BC019411 Mus muscu
11	1174	75.9	178965	9	AC098850 Homo sapi
12	1174	75.9	192337	9	AC022596 Homo sapi
13	1165	75.4	211403	9	AC124352 Homo sapi
14	1150.5	74.4	207418	9	AC090774 Homo sapi
15	1145.5	74.1	176040	9	AC107926 Homo sapi
16	1114	72.1	188017	2	AC146897 Gorilla g
17	1016	65.7	131239	9	HS667H12 Human DNA
18	956	61.8	2024	9	AY063502 Cercopith
19	923	59.7	1992	9	AY063503 Macaca as
20	882	57.1	1293	5	CHK8STFL25 D26340 Gallus gall
21	816	52.8	2165	9	AF448495 Callithri
22	755.5	48.9	1911	9	AK057473 Homo sapi
23	755.5	48.9	3459	9	BC041488 Homo sapi
24	423.5	27.4	1153	3	AY089374 Drosophil
25	375	24.3	1460	3	AF032921 Drosophil
26	373.5	24.2	1117	3	TBU74470 Trypanosoma
27	373.5	24.2	141775	2	AC105378 Trypanoso
28	356.5	23.1	37983	2	AC020394 Drosophil
29	356.5	23.1	85651	2	AC005463 Drosophil
30	356.5	23.1	86889	3	AC005448 Drosophil
31	356.5	23.1	139410	3	AC007084 Drosophil
32	356.5	23.1	254023	3	AE003839 Drosophil
33	331.5	21.4	34828	3	U41994 Caenorhabdi
34	312	20.2	166143	2	AL356960 Homo sapi
35	303.5	19.6	2277	3	TRFRNH1A f
36	302	19.5	43680	2	AC145667 Homo sapi
37	300.5	19.4	56083	9	AC108488 Homo sapi
38	292.5	18.9	38360	2	AC145674 Homo sapi
39	286	18.5	206332	10	AC136986 Mus muscu
40	276	17.9	39512	8	SPBC336 A121815 S.pombe c
41	272	17.6	1557	8	AF048992 Schizosac
42	272	17.6	119211	6	AK408761 Sequence
43	272	17.6	119211	6	AX067465 Sequence
44	268	17.3	231	6	AR422883 Sequence
45	268	17.3	231	6	BD118436 EST and e

ALIGNMENTS

RESULT 1

Tue Mar 23 14:48:35 2004

BC002973	1158 bp	mRNA	linear	PRI 12-NOV-2003
LOCUS				
DEFINITION				Homo sapiens ribonuclease H1, mRNA (CDNA clone MGC:2019 IMAGE:3537074), complete cds.
ACCESSION				BC002973
VERSION				BC002973.1
KEYWORDS				GI:12804228
SOURCE				MGC.
ORGANISM				Homo sapiens (human)
REFERENCE				
AUTHORS				1 (bases 1 to 1158) Strausberg,R.L., Peingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,J., Stenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marasina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheet,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butlerfield,Y.S., Krzyzinski,M.I., Skalska,U., Smalius,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE				Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL				Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE				22388257
PUBMED				12477932
REFERENCE				2 (bases 1 to 1158) Strausberg,R.
AUTHORS				Direct Submission
JOURNAL				Submitted (06-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550, USA
REMARK				NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>
COMMENT				Contact: MGC help desk Email: <a href="mailto:cgabbs-r@mail.nih.gov">cgabbs-r@mail.nih.gov</a> Tissue Procurement: DCTB/Drp CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada <a href="mailto:info@bcgsc.bc.ca">info@bcgsc.bc.ca</a> Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fiehl, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzyzinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalius, Michael Smith, Lorraine Spence, Jeff Scott, Michael Thorne, Miranada Teai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.
FEATURES				Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Series: IRAL Plate: 8 Row: 1 Column: 11 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21359815. Location/Qualifiers 1..1158 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="MGC:2019 IMAGE:3537074" /tissue type="Lung, small cell carcinoma" /clone_tib="vntH MGC 7"

```
QY 181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200
DB 611 ACAACCAAGAGCGGAATTCATGCGAGCTGCAAGCAATGAAACCAAGCAAGACTCAA 670
QY 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
DB 671 AACATCAATAAATCGTCTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC 730
QY 221 TrpValGlnGlyTyrLysLysAsnGlyTyrLysThrSerAlaGlyLysGluValIleAsn 240
DB 731 TGGGTTCAGGTTGGAAGAAAATGGGTGGAAGACAAGTGCAGGGAAGAGTGTATCAAC 790
QY 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTyrMetHis 260
DB 791 AAAGAGGACTTGTGCTGCTGAGAGAGGCTTACCCAGGAGTGCATCAATTCAGTGGATGCAT 850
QY 261 ValProGlyHisSerGlyPheIleGlyAsnGluAlaAspArgLeuAlaArgGluGly 280
DB 851 GTTCTCGTGCATTCGGGATTTATAGCAATGGAAGCTGACAGATTAGCCAGAGAGGA 910
QY 281 AlaLysGlnSerGluAsp 286
DB 911 GCTAAACAATCGAAGAC 928

RESULT 2
AF039652 1168 bp mRNA linear PRI 02-APR-1998
LOCUS Homo sapiens ribonuclease H type II mRNA, complete cds.
DEFINITION AF039652
ACCESSION AF039652.1 GI:3004980
VERSION
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1168)
AUTHORS Wu, H., Lima, W. and Crooke, S.
TITLE Molecular cloning and expression of cDNA for human RNase H
JOURNAL Antisense Nucleic Acid Drug Dev. (1998) in press
REFERENCE 2 (bases 1 to 1168)
AUTHORS Wu, H., Lima, W. and Crooke, S.
TITLE Direct Submission
JOURNAL Submitted (22-DEC-1997) Molecular Pharmacology, Isis Pharmaceuticals, Inc., 2292 Faraday Ave, Carlsbad, CA 92008, USA

FEATURES
Location/Qualifiers
1..1168
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17p11.2"
82..942
/function="ribonuclease specific for the RNA strand of a DNA-RNA hybrid"
/notes="RNAase HII"
/codon_start=1
/product="ribonuclease H type II"
/protein_id="AAC09261.1"
/db_xref="GI:3004981"
/translation="MSLLSLAHEVALAALPCRRSGFGMGFVAVRGRKTVGLTWN
ECRAQVDRFAARFKFATDEAWAFVRKASPEVSEGHENQHQSEAKSLREP
LDGDSAPFPAYAKMPKSPVEPPARDTFSYMGDFVVTIDCCSSNGRRPRAGI
GVYWGPGHNLVGRLEQRTNQRAIEHAACKAEQAKTNINKLVLYTDSMTINGI
TNVQWKKNGKWTSGKVEINKEDFVALERLTQGMIDIQWHPVGHSGFTIGNEADRL
AREGAKQSED"

CDS
82..942
/function="ribonuclease specific for the RNA strand of a DNA-RNA hybrid"
/notes="RNAase HII"
/codon_start=1
/product="ribonuclease H type II"
/protein_id="AAC09261.1"
/db_xref="GI:3004981"
/translation="MSLLSLAHEVALAALPCRRSGFGMGFVAVRGRKTVGLTWN
ECRAQVDRFAARFKFATDEAWAFVRKASPEVSEGHENQHQSEAKSLREP
LDGDSAPFPAYAKMPKSPVEPPARDTFSYMGDFVVTIDCCSSNGRRPRAGI
GVYWGPGHNLVGRLEQRTNQRAIEHAACKAEQAKTNINKLVLYTDSMTINGI
TNVQWKKNGKWTSGKVEINKEDFVALERLTQGMIDIQWHPVGHSGFTIGNEADRL
AREGAKQSED"

ORIGIN
Alignment Scores: 1.73e-101 Length: 1168
Pred. No.: 1526.00 Matches: 282
Score: 99.30% Conservative: 2
Percent Similarity: 98.60% Mismatches: 2
Best Local Similarity: 98.60%
```

```
Query Match: 98.71% Indels: 0
DB: 9 Gaps: 0
US-10-054-313-1 (1-286) x AF039652 (1-1168)
QY 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
DB 82 ATGAGCTGGCTTCTGTTCTGCGCCACAGAGTCGCTTGGCGCCCTTGGCTGCCCGCG 141
QY 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40
DB 142 GGCTCTCGCGGTTCGGGATGTTCTATGCGGTGAGGAGGGCGCCGAGACCCGGGTCTTT 201
QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
DB 202 CTGACCTGGGAATGAGTGCAGAGCAGAGTGGACCGGTTCCTGCTGCCAGATTAAGAAG 261
QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
DB 262 TTTGCCACAGAGGATGAGGCTGGGCTTTGTCTGAGGAATCTGCAAGCCCGAAGTTTCA 321
QY 81 GluGlyHisGluAsnGlnHisGlyGlnLysGluAlaLysProGlyLysArgLeuArg 100
DB 322 GAAGGGCATGAAATCAACATGACACAGAATCGAGGCGGAAGCCAGCAAGCGACTCCGT 381
QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
DB 382 GAGCCACTGGATGGAGATGACATGAAAGCGCAGAGCCGTATGCAAGACACATGAAGCCG 441
QY 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
DB 442 AGCGTGGAGCGCGCCCTCCAGTTAGCAGAGACACGTTTCTTCATCGGAGACTTGGTC 501
QY 141 ValValTyrThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGlyLe 160
DB 502 GTGCTCTACACTGATGCTGCTGCTCCAGTAATGGCGGTAGAGGCGCGCAGCAGGAATC 561
QY 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyLysLeuArgLeuProGlyArgGln 180
DB 562 GCGCTTACTGGGGGCCAGCCATCCCTTTAAATGTAGCATATAGACTTCCTGGCGCGGAG 621
QY 181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200
DB 622 ACAACCAAGAGCGGAATTCATGCGAGCTGCAAGCAATGAAACCAAGCAAGACTCAA 681
QY 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
DB 682 AACATCAATAAATCGTCTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC 741
QY 221 TrpValGlnGlyTyrLysLysAsnGlyTyrLysThrSerAlaGlyLysGluValIleAsn 240
DB 742 TGGGTTCAGGTTGGAAGAAAATGGGTGGAAGACAAGTGCAGGGAAGAGTGTATCAAC 801
QY 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTyrMetHis 260
DB 802 AAAGAGGACTTGTGCTGCTGAGAGGCTTACCCAGGAGTGCATTCAGTGGATGCAT 861
QY 261 ValProGlyHisSerGlyPheIleGlyAsnGluAlaAspArgLeuAlaArgGluGly 280
DB 862 GTTCTCGTGCATTCGGGATTTATAGCAATGGAAGCTGACAGATTAGCCAGAGAGGA 921
QY 281 AlaLysGlnSerGluAsp 286
DB 922 GCTAAACAATCGAAGAC 939

RESULT 3
AK075490 1604 bp mRNA linear PRI 03-SEP-2002
LOCUS Homo sapiens cDNA PSEC0185 fis, clone OVARC1002091, highly similar
DEFINITION to Homo sapiens ribonuclease H type II mRNA.
ACCESSION AK075490
VERSION AK075490.1 GI:22761667
KEYWORDS oligo capping; fis (full insert sequence).
```

SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
 1 Ota, T., Nishikawa, T., Suzuki, Y., Kawai-Hio, Y., Hayashi, K.,  
 Ishii, S., Saito, K., Yamamoto, J., Makamatsu, A., Nagai, T.,  
 Nakamura, Y., Nagahara, K., Sugano, S. and Isogai, T.  
 HRI human cDNA sequencing project  
 Unpublished  
 2 (bases 1 to 1604)  
 Isogai, T. and Yamamoto, J.  
 Direct Submission  
 Submitted (20-MAR-2002) Takao Isogai, Helix Research Institute,  
 Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

TITLE  
 JOURNAL  
 HRI human cDNA sequencing project; cDNA 5'- & 3'-end one pass  
 sequencing, clone selection and full insert sequencing; Helix  
 Research Institute (supported by Japan Key Technology Center etc.);  
 cDNA library construction; Institute of Medical Science, University  
 of Tokyo, Laboratory of Genome Structure, Human Genome Center.

FEATURES  
 Location/Qualifiers  
 1..1604  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="OVARC1002091"  
 /tissue\_type="ovary, tumor tissue"  
 /clone\_lib="OVARC1"  
 /notes="cloning vector: pME18SFL3"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 2,46e-101 Length: 1604  
 Score: 1526.00 Matches: 282  
 Percent Similarity: 99.30% Conservative: 2  
 Best Local Similarity: 98.60% Mismatches: 2  
 Query Match: 98.71% Indels: 0  
 DB: 9 Gaps: 0

US-10-054-313-1 (1-286) x AK075490 (1-1604)

Qy 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20  
 Db 61 ATGAGCTGGCTTCGTCTCTGGCCACAGAGTCGCTTGCCGCGCTTGCCTCCGCGC 120  
 Qy 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyValGlyValPhe 40  
 Db 121 GGCTCTCGCGGTTCCGGATTTCTATGCCGTGAGGAGGGCCGCGCAGACCGGGCTTT 180  
 Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLys 60  
 Db 181 CTGACCTGGATGATGTCAGAGCACAGGTGACCGCTTCTCTGTCGACGATTTAAGAAG 240  
 Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80  
 Db 241 TTTCGCCACAGAGGATGAGCTTGGCCCTTTCTCAGGAAATCTGCAAGCCCGAAGTTTCA 300  
 Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100  
 Db 301 GAGGGCATGAAATCAACATGGACACAGATCGGAGCGGAAAGCCAGCAGCGACTCCGT 360  
 Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120  
 Db 361 GAGCCACTGGATGGAGATGGACATGAAAGCCGACGCGTATGCAAGACATGAAGCCG 420  
 Qy 121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140  
 Db 421 ACCGTGGAGCCGCGCCCTCCAGTTAGCAGAGACACGTTTCTCATCGGAGACTTCGTC 480  
 Qy 141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyTle 160  
 Db 481 GTTCGTCTACTGATGGCTGCTGCTCCAGTATGGCGGTAGAAAGCCCGCAGCGGAATC 540

Qy 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyTleArgLeuProGlyArgGln 180  
 Db 541 GCGGTTTACTGGGGCGGCGCCATCTCTTAAATGTAGGCATTAGACTTCTCTGGCGCAG 600

Qy 181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200  
 Db 601 ACAACCCAAAGAGCGGGAATTCATGCAGCGCTGCAAGCCATTCAACAGCAAGACTCAA 660

Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyTleThrAsn 220  
 Db 661 AACATCAATAAATCTGTTCTGTATACAGACAGTATGTTACGATAAATGTTAATCAAC 720

Qy 221 TrpValGlnGlyTrpLysLysAsnGlyTyrLysThrSerAlaGlyLysGluValIleAsn 240  
 Db 721 TGGGTTTCAAGTTGGAGAAATATGGTGGACAGCAAGTCAGGAAAGAGGTGATCAAC 780

Qy 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTyrMetHis 260  
 Db 781 AAAGAGGACTTTGTGGCACTGGAGAGCTTACCCAGGGGATGGACATTCAGTGGATGCAT 840

Qy 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280  
 Db 841 GTTCCTGCTCATTCGGGATTTATAGGCATGAGAGAGCTGACAGATTAGCCAGAGAAGGA 900

Qy 281 AlalysGlnSerGluAsp 286  
 Db 901 GCTAAACAATCGGAAGAC 918

RESULT 4  
 BD223780 1150 bp DNA linear PAT 17-JUL-2003  
 LOCUS Human RNA-associated proteins.  
 DEFINITION BD223780  
 ACCESSION BD223780.1 GI:33033550  
 VERSION JP 2002523045-A/16.  
 KEYWORDS Homo sapiens (human)  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1150)  
 Hillman, J.L., Yue, H., Tang, Y.T., Corley, N.C., Guegler, K.J.,  
 Gorgone, G.A., Patterson, C., Baughn, M.R., Lal, P., Bandman, O.,  
 Reddy, R., Azimzai, Y., Shih, L.L., Yang, J. and Lu, D.A.M.  
 Human RNA-associated proteins  
 Patent: JP 2002523045-A 16 30-JUL-2002;  
 INCYTE PHARMACEUTICALS INC  
 OS Homo sapiens (human)  
 PN JP 2002523045-A/16  
 PD 30-JUL-2002  
 PF 20-AUG-1999 JP 2000566425  
 PR 21-AUG-1998 US 60/097550, 12-JAN-1999 US 60/115639 PI  
 JENNIFER L HILLMAN, HENRY YUE, Y TOM TANG, NEIL C CORLEY, KARL J PI  
 GUEGLER,  
 PI GINA A GORGONE, CHANDRA PATTERSON, MARIAH R BAUGHN, PREETI LAL,  
 PI OLGA BANDMAN,  
 PI ROOPA REDDY, YALDA AZIMZAI, LEO L SHIH, JUNMING  
 YANG, DYUNG AINA M  
 PI LU  
 PC C12N15/09, A61K38/00, A61K38/55, A61K45/00, A61P1/00, A61P1/04, PC  
 A61P1/16,  
 PC A61P1/18, A61P3/00, A61P3/10, A61P7/00, A61P7/06, A61P9/00, A61P11/  
 00, A61P13/06,  
 PC A61P13/12, A61P17/00, A61P17/06, A61P19/00, A61P19/02, A61P19/06,  
 PC A61P19/10,  
 PC A61P21/00, A61P21/04, A61P29/00, A61P29/00, A61P31/00, A61P31/04,  
 PC A61P31/10,  
 PC A61P31/12, A61P31/18, A61P31/22, A61P35/00, A61P35/02, A61P37/00,  
 PC A61P37/08,  
 PC C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/10 PC  
 C12P21/02, C12P1/68,  
 PC G01N33/15, G01N33/50, G01N33/53, G01N33/566//C12P21/08, C12N15/00,  
 PC C12N5/00,

PC A61K37/02.A61K37/64  
 CC Incyte Identification No.: 2073417CB1  
 FH Key Location/Qualifiers  
 FT source 1..1150  
 FT /organism="Homo sapiens (human)"

## FEATURES

Location/Qualifiers  
 1..1150  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

## ORIGIN

Alignment Scores:  
 Pred. No.: 2,88-101 Length: 1150  
 Score: 1523.00 Matches: 281  
 Percent Similarity: 99.30% Conservatives: 3  
 Best Local Similarity: 98.25% Mismatches: 2  
 Query Match: 98.51% Indels: 0  
 DB: 6 Gaps: 0

US-10-054-313-1 (1-286) x BD223780 (1-1150)

QY 1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20  
 DB 85 ATGAGCTGGCTTCTGTTCTGCCCCACAGAGTCGCTTGGCGCGCTTGCCTGCGCGCGC 144  
 QY 21 GlySerArgGlyPheGlyMetPheTrpAlaValArgGlyArgGlyThrGlyValPhe 40  
 DB 145 GGCTCTCGCGGTTCTGGGATGTTCTATCGGTGAGAGGGCGCCGACACCGGGTCTTT 204  
 QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60  
 DB 205 CTGACCTGGAATGAGTCAGACAGACACAGGTGGACCGTTTCTGCTGCCAGATTTAAGAAG 264  
 QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80  
 DB 265 TTTGCCACAGAGGATGAGCGCTTGGCGCTTGTTCAGGAATCTGCAAGCCCGGAAGTTTCA 324  
 QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100  
 DB 325 GAAGGCAATGAATCAATGACACAGATCGGAGCGGAAGCCAGCAGCGACTCCGT 384  
 QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyAlaLysHisMetLysPro 120  
 DB 385 GAGCCACTGGATGGAGATGACATGAAGCGCAGACCGGTATGCAAGCACATGAAGCGC 444  
 QY 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140  
 DB 445 AGCATGGAGCGCGCTTCAGTTAGCAGACACGCTTTTCTACATGGGAGACTTCGTC 504  
 QY 141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyLys 160  
 DB 505 GTCGTCTACACTGATGGCTGCTGCTCAGTAATGGCGGTAGAAGCGCGCGAGCAATC 564  
 QY 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyLysArgLeuProGlyArgGln 180  
 DB 565 GCGTTTACTGGGGCGAGCCCATCTTTAAATGTAGGATATTAGACTTCTGGCGGCGAG 624  
 QY 181 ThrAsnGlnArgAlaGluLeuHisAlaAlaCysLysAlaLeuGluGlnAlaLysThrGln 200  
 DB 625 ACAACCAAGAGCGGAATTCATGCGCGCTGCAAGCCATTGAACAAGCAAGACTCAA 684  
 QY 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyLysThrAsn 220  
 DB 685 AACATCAATAAATGGTTCTGTATACAGACAGATGTTTACGATAAATGGTATACTAAC 744  
 QY 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240  
 DB 745 TGGTTCAAGGTTGGAAGAAAATCGTGTGAAGCAAGTCAGCGAAGAGGTGATCAAC 804  
 QY 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260  
 DB 805 AAAGAGGACTTGTGGCACTGGAGAGGCTTACCAGGGGATGACATTCAGTGGATGCAT 864

QY 261 ValProGlyHisSerGlyPheIleGlyVaenGluAlaAspArgLeuAlaArgGluGly 280  
 DB 865 GTTCTCTGGTCATTCGGGATTTATAGCAATGAGAGCTGACAGATTAGCCAGAGAGGA 924  
 QY 281 AlaLysGlnSerGluAsp 286  
 DB 925 GCTAAACAATCGGAGAC 942

## RESULT 5

AF048994 1113 bp mRNA linear PRI 21-NOV-1998  
 LOCUS Homo sapiens ATCC158373 ribonuclease H1 (rnhl) mRNA, complete cds.  
 ACCESSION AF048994  
 VERSION AF048994.1 GI:2935441  
 KEYWORDS  
 SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 1113)  
 Cerretti,S.M., Fedoroff,O.Y., Reid,B.R. and Crouch,R.J.  
 A common forty amino acid motif in eukaryotic RNases H1 and  
 mouse related to bacterial RNase H1  
 Nucleic Acids Res. (1998) In press

## JOURNAL

2 (bases 1 to 1113)  
 Cerretti,S.M. and Crouch,R.J.  
 Cloning, expression, and mapping of ribonucleases H of human and  
 mouse related to bacterial RNase H1  
 Genomics 53 (3), 300-307 (1998)

## AUTHORS

99017966  
 9799596  
 3 (bases 1 to 1113)  
 Cerretti,S.M. and Crouch,R.J.  
 Direct Submission

## TITLE

Submitted (17-FEB-1998) Laboratory of Molecular Genetics, NICHD  
 NIH, 6 Center Drive Building 6B Room 2B-231, Bethesda, MD 20892,  
 USA

## FEATURES

## source

Location/Qualifiers  
 1..1113  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /strain="ATCC158373"  
 /db\_xref="ATCC:158373"  
 /db\_xref="dbEST:AA356084"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="17p11.2"  
 /cell\_type="Jurkat T-cells VI"

## gene

1..1113  
 /gene="rnhl"

## CDS

28..888  
 /gene="rnhl"  
 /note="similar to bacterial RNases H1"  
 /codon\_start=1  
 /product="ribonuclease H1"  
 /protein\_id="AAC78563.1"  
 /db\_xref="GI:2935442"  
 /translation="MSWFLFAHRVALAALPCRGSRGFGMFAVRRGRKTVGLTWN  
 ECAQVDRFPFAFKFATDEANAVFKASPEVSEGHENQHQESSEAKSKRLREP  
 LDGSHSAPFAAKHKPSPVEAPPVSRDFTSMGDFVYVYTTGCCSSNGRRPRAGI  
 GVYWGPGHNVGIRLPGRTNQRIEHAACKAIEQAKTQNKILVYTDSTMTFINGI  
 TNYVQGNKQGNKTSAGKEVINKEDFVALERLTQGMIDIQWHEVGHSGFTIGNEADRL  
 AREGAKQSED"

## ORIGIN

Alignment Scores:  
 Pred. No.: 3,19e-101 Length: 1113  
 Score: 1522.00 Matches: 281  
 Percent Similarity: 98.95% Conservatives: 2  
 Best Local Similarity: 98.25% Mismatches: 3  
 Query Match: 98.45% Indels: 0  
 DB: 9 Gaps: 0

US-10-054-313-1 (1-286) x AF048994 (1-1113)

QY 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20  
 Db 28 ATGAGCTGTTCTTCTTCTGCCCACAGTCGCTTGGCCGCTTGGCCCTGCCGCGCGC 87  
 QY 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40  
 Db 88 GGCCTCTCGGGGTTCCGGATGTTCTATGCCGTGAGAGGGCGCGCAAGACCGGGTCTTT 147  
 QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspGlyPheProAlaAlaArgPheLysLys 60  
 Db 148 CTGACCTGGATGATGTCGACAGACAGTCGGACCGGTTTCTTCTGCGCCAGATTAAAGAG 207  
 QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80  
 Db 208 TTGTGCACACAGAGATGAGCGCTTGGCCCTTTGTTCAGGAATCTGCAAGCCGCGAAGTTTCA 267  
 QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100  
 Db 268 GAAGGCGCATGAAATCAATGATGACAGAAATCGAGCGCAAGCCGACGAGCATCTCGT 327  
 QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120  
 Db 328 GAGCCACTGGATGGATGACATGATGAAGCGCAGCGCGTATGCAAGCACATGAAGCGC 387  
 QY 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140  
 Db 388 AGCGTGGAGCGCGCGCTTCAGTTAGCAGACACCGTTTCTTACATGGGAGCTTCGTC 447  
 QY 141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160  
 Db 448 GTCGTCTACACTGATGGCTGCTCTCCAGTAAATGGCGTAGAGCGCGGACGAGGATC 507  
 QY 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180  
 Db 508 GCGCTTTACTGGGGCGCGAGCCATCTCTTAAATGTAGGATAGACTTCTCTGGCGCGCAG 567  
 QY 181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200  
 Db 568 ACACACCAAGACGGGAATTCATGACGCTCGCAAGCCATTGAACACCAAGACTCAA 527  
 QY 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220  
 Db 628 AACATCAATAAATGCTTCTGTATACAGACAGTATGTTACGATAAATGGTATACTAAC 687  
 QY 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240  
 Db 688 TGGGTTCAAGGTTGGAAGAAAATGGGTGGAAGACAAAGTGCAGGGAAGAGGTGATCAAC 747  
 QY 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260  
 Db 748 AAGAGGACCTTTGTGGCACTGGAGAGGCTTACCAGGGGATGACATTCAGTGGATGCAT 807  
 QY 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280  
 Db 808 GTTCTCTGCTATTGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGGA 867  
 QY 281 AlalysGlnSerGluAsp 286  
 Db 868 GCTAAACAAATCGGAGAC 885

## RESULT 6

AF048995 1147 bp mRNA linear PRI 21-NOV-1998  
 LOCUS Homo sapiens ATCC159806 ribonuclease H1 (rnhi) mRNA, complete cds.  
 DEFINITION  
 ACCESSION AF048995  
 VERSION AF048995.1 GI:2935443  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## REFERENCE

AUTHORS Carricelli,S.M., Fedoroff,O.Y., Reid,B.R. and Crouch,R.J.  
 TITLE A common forty amino acid motif in eukaryotic RNases H1 and caulimovirus ORF VI proteins binds to duplex RNAs  
 JOURNAL Nucleic Acids Res. (1998) In press  
 REFERENCE 2 (bases 1 to 1147)  
 AUTHORS Carricelli,S.M. and Crouch,R.J.  
 TITLE Cloning, expression, and mapping of ribonucleases H of human and mouse related to bacterial RNase H1  
 JOURNAL Genomics 53 (3), 300-307 (1998)  
 MEDLINE 99017966  
 PUBMED 979596  
 REFERENCE 3 (bases 1 to 1147)  
 AUTHORS Carricelli,S.M. and Crouch,R.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-FEB-1998) Laboratory of Molecular Genetics, NICHD NIH, 6 Center Drive Building 6B Room 2B-231, Bethesda, MD 20892, USA

## FEATURES

## source

## Location/Qualifiers

1..1147  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /strain="ATCC159806"  
 /db\_xref="ATCC:159806"  
 /db\_xref="dbEST:AA312554"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="17p11.2"  
 /cell\_type="Jurkat T-cells VI"  
 1..1147  
 /gene="rnhi"  
 62..922  
 /gene="rnhi"  
 /note="similar to bacterial RNases H1"  
 /codon\_start=1  
 /product="ribonuclease H1"  
 /protein\_id="AAC78564.1"  
 /db\_xref="GI:2935444"  
 /translation="MSWFLAHRVALAALPCRSGFGMFAVRRGRTGTVLTVN  
 ECRQVRFPAARFKKATEDEAWFVKASPEVSEGHENQHQSEAKASRLREP  
 LDGQHSASPYAKMKPSPVEAPPVSDTFSYMGDFVVVYTDCCSSNGRRRPRAGI  
 GYVGPQHPNLVNLGRLPGRQTNQRAEIHAAKALEQAKTONKLVLYTDSMTINGI  
 TNWQGWKKNGWKTSAGKEINKEDFVALERLTQGMDIQMHVPHSGFGFIGNBEADRL  
 AREGAKQSED"

## gene

## CDS

## ORIGIN

## Alignment Scores:

Prod. No.: 3,3e-101 Length: 1147  
 Score: 1522.00 Matches: 281  
 Percent Similarity: 98.95% Conservatives: 2  
 Best local Similarity: 98.25% Mismatches: 3  
 Query Match: 98.45% Indels: 0  
 DB: 9 Gaps: 0

US-10-054-313-1 (1-286) x AF048995 (1-1147)

QY 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20  
 Db 62 ATGAGCTGTTCTTCTTCTGCCCACAGATCGCCTTGGCCGCTTGGCCCTGCCGCGC 121  
 QY 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40  
 Db 122 GGCCTCTCGGGGTTCCGGATGTTCTATGCCGTGAGAGGGCGCGCAAGACCGGGTCTTT 181  
 QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60  
 Db 182 CTGACCTGGAATGATGTCGACAGACACAGGTGACCGGTTTCTGCTGCCAGATTAAAGAG 241  
 QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80  
 Db 242 TTGTCCACAGAGATGAGGCTTGGCCCTTTGTTCAGGAATCTGACGCCCGGAAAGTTTCA 301

Mammalia; Rutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 1147)

Carricelli,S.M., Fedoroff,O.Y., Reid,B.R. and Crouch,R.J.

A common forty amino acid motif in eukaryotic RNases H1 and caulimovirus ORF VI proteins binds to duplex RNAs

Nucleic Acids Res. (1998) In press

2 (bases 1 to 1147)

Carricelli,S.M. and Crouch,R.J.

Cloning, expression, and mapping of ribonucleases H of human and mouse related to bacterial RNase H1

Genomics 53 (3), 300-307 (1998)

99017966

979596

3 (bases 1 to 1147)

Carricelli,S.M. and Crouch,R.J.

Direct Submission

Submitted (17-FEB-1998) Laboratory of Molecular Genetics, NICHD NIH, 6 Center Drive Building 6B Room 2B-231, Bethesda, MD 20892, USA

```
Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db 302 GAAGGGCATGAAATCAACATGCAAGATCGGAGCGAAGCCAGCAAGCGATCCGT 361
Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db 362 GAGCCACTGGATGGAGATGGACATGAAGCGCAGAGCGGTATGCAAGCACATGAAGCG 421
Qy 121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
Db 422 AGCGTGGAGCGCGCCCTCCAGTTAGCAGACACAGTTTCTTCTACATGGAGACTTCGTC 481
Qy 141 ValValTyrThrAspGlyCysSerSerAsnGlyArgGlyLysProArgAlaGlyIle 160
Db 482 GTCGCTCACATGATGGCTGCTCCATGATGGCGGTATGAGGCCCGCGAGCAGGATC 541
Qy 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
Db 542 GCGGTTTACTGGGGCCAGGCCATCTTTAAATGTAGGCATTAGACTTCTCTGGCGCGAG 601
Qy 181 ThrAsnGlnArgAlaGluIleHisAlaLysAlaLysAlaLysGluGlnAlaLysThrGln 200
Db 602 ACAACCAAGAGCGGAATTCATGCGCTGCAAGCCATTGAACAAGCAAGCAAGACTCAA 661
Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db 662 AACATCAATAAATGGTTCTGTATACAGACATGTTTACGATAAATGGTATAACTAAC 721
Qy 221 TrpValGlnGlyTyrLysAsnGlyTyrLysThrSerAlaGlyLysGluValIleAsn 240
Db 722 TCGGTTCAAGTGTGAAGAAATCGGTGGAAGCAAGTGCAGGGAAGAGGTGATCAAC 781
Qy 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Db 782 AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAGTGGATGAT 841
Qy 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280
Db 842 GTTCTGTGTCATTTCGGATTTATAGCAATGAAGAAGCTGACAGATTAGCCAGAGAGGA 901
Qy 281 AlaLysGlnSerGluAsp 286
Db 902 GCTAACAAATCGAAGAC 919
```

## RESULT 7

```
HSAJ4117
LOCUS HSAJ4117 861 bp mRNA linear PRI 07-APR-1998
DEFINITION Homo sapiens mRNA RNase HII.
ACCESSION AJ224117
VERSION AJ224117.1 GI:3043446
KEYWORDS RNase HII; rnhII gene.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 861)
AUTHORS Frank, P., Braunschofer-Reiter, C. and Wintersberger, U.
TITLE Cloning and functional expression of human RNase HII
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 861)
AUTHORS Frank, P., Braunschofer-Reiter, C. and Wintersberger, U.
TITLE Direct Submision
JOURNAL Submitted (22-FEB-1998) Department of Molecular Genetics, Institute of Tumor Biology-Cancer Research, Borschkegasse 8a, Vienna, Austria A-1090
```

## FEATURES

## source

```
Location/Qualifiers
1..861
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..861
/genes="rnhII"
1..861
```

## gene

## CDS

```
/genes="rnhII"
/notes="evolutionarily related to prokaryotic RNase HI and
RNase H domain of retroviral reverse transcriptase;
ribonuclease HII"
/codon_start=1
/product="RNase HII"
/protein_id="CAAL1835.1"
/db_xref="GI:3043447"
/db_xref="SWISS-PROT:O60930"
/translation="MSWLFLAHRVALAALPCRRSGRGFMFYAVRRGRKTGVFLTNW
ECRAQVDRFPAARFKFAFEDAEAWAFKASASPEVSEGHENHGRSESAKAKRRREP
LDGDGESAEPAKHKPSVPAPVSRDTFPMGDFVVVTDGCSNGRRRRPAGI
GYWGPCHEPLNIGIRLPGRTNORAEIHAACKALQAKTNKLVLYDTSFTINGI
TNVRGKKNKGNWTSAGKEVINKEDFVALERLTQMDIQMHVPHSGFIGNEDRL
AREGAKQSED"
```

## ORIGIN

```
Alignment Scores:
Pred. No.: 4,66e-101 Length: 861
Score: 1518.00 Matches: 280
Percent Similarity: 99.30% Conservative: 4
Best Local Similarity: 97.90% Mismatches: 2
Query Match: 98.19% Indels: 0
DB: 9 Gaps: 0
US-10-054-313-1 (1-286) x HSAJ4117 (1-861)
Qy 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db 1 ATGAGCTGGCTTCTGTCTCGGCCACAGAGTCGCTTGGCCCTTGGCCCTGCCGCCGC 60
Qy 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyArgLysThrGlyValPhe 40
Db 61 GCTCTCGCGGTTCGGGATGTTCTATGCGTGAGAGGGGGCGCAAGCCGGGTCTTT 120
Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaLysPheLysLys 60
Db 121 CTGACCTGGAAATGAGTGCAGAGCACAGTGGACCGGTTTCTCTCTCCAGATTAAAG 180
Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 181 TTTGCCACAGAGATGAGGCTTGGGCTTTGTGAGGAAATCTGCACAGCCCGAGATTCA 240
Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db 241 GAAGGGCATGAAATCAACATGACGAGAAATCGAGCGAAGCCAGCAAGCGACTCCGT 300
Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db 301 GAGCCACTGGATGGAGATGGACATGAAGCGCAGAGCGCGTATGCAAGCACATGAAGCCG 360
Qy 121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
Db 361 AGCGTGGAGCGCGCCCTCCAGTTAGCAGACACAGTTTCTTCTACATGGAGACTTCGTC 420
Qy 141 ValValTyrThrAspGlyCysSerSerAsnGlyArgGlyLysProArgAlaGlyIle 160
Db 421 GTCGCTCACATGATGGCTGCTCCATGATGGCGGTATGAGGCCCGCGAGCAGGATC 480
Qy 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
Db 481 GCGGTTTACTGGGGCCAGGCCATCTTTAAATGTAGGCATTAGACTTCTCTGGCGCGAG 540
Qy 181 ThrAsnGlnArgAlaGluIleHisAlaLysAlaLysAlaLysGluGlnAlaLysThrGln 200
Db 541 ACAACCAAGAGCGGAATTCATGCGCTGCAAGCCATTGAACAAGCAAGCAAGACTCAA 600
Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db 601 AACATCAATAAATGGTTCTGTATACAGACATGTTTACGATAAATGGTATAACTAAC 660
Qy 221 TrpValGlnGlyTyrLysLysAsnGlyTyrLysThrSerAlaGlyLysGluValIleAsn 240
```



```

|||||...|||||
661 TCGGTCGAGGTTGGAAGAAATCGGTGGAAGACAAGTCGAGGAAGAGGTATCAAC 720
Db

241 LysGluAspPheValAlaLeuGluAArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Qy

721 AAAGAGGACTTTGTGGCACTCGAGAGGCTTACCAGGGGATGGACATTCAGTGGATGCAT 780
Db

261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAAspArgLeuAlaArgGluGly 280
Qy

781 GTTCTCGGTTCATTCGGGATTTATAGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGGA 840
Db

281 AlaLysGlnSerGluAsp 286
Qy

841 GCTAACAAATCGGAGAC 858
Db

RESULT 8
AK096913          2120 bp mRNA linear PRI 15-JUL-2002
LOCUS             Homo sapiens cDNA FLJ39594 fis, clone SKNSH2001875, highly similar
DEFINITION        to Homo sapiens ATCC158373 ribonuclease H1 (rnhl) mRNA.
ACCESSION         AK096913
VERSION           AK096913.1 GI:21756512
KEYWORDS           oligo capping; fis (full insert sequence).
SOURCE            Homo sapiens
ORGANISM          Homo sapiens
REFERENCE         1
AUTHORS           Kawakami,B., Sugiyama,A., Takemoto,M., Sugiyama,T., Irie,R.,
                  Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
                  Kawai-Hio,Y., Saito,K., Nishikawa,F., Kimura,K., Yamashita,H.,
                  Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
                  Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
                  Oshima,A., Suzuki,Y., Sugano,S., Nagahara,K., Masuho,Y., Nagai,K.
                  and Isogai,T.
TITLE             NEDO human cDNA sequencing project
JOURNAL           Unpublished
REFERENCE         2 (bases 1 to 2120)
AUTHORS           Isogai,T. and Yamamoto,J.
TITLE             Direct Submission
JOURNAL           Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
                  Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
                  (E-mail:genomics@hri.co.jp, Tel.81-438-52-3975, Fax:81-438-52-3986)
COMMENT           NEDO human cDNA sequencing project supported by Ministry of
                  Economy, Trade and Industry of Japan; cDNA full insert sequencing:
                  Research Association for Biotechnology (RAB); cDNA library
                  construction: Helix Research Institute (HRI) (supported by Japan
                  Key Technology Center etc.); 5' - 3'-end one pass sequencing: RAB,
                  HRI, and Biotechnology Center, National Institute of Technology and
                  Evaluation; clone selection for full insert sequencing: HRI and
                  RAB; annotation: HRI and RAB.
FEATURES          source
                  1..2120
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="SKNSH2001875"
                  /cell_line="SK-N-SH"
                  /cell_type="neuroblastoma"
                  /clone_lib="SKNSH2"
                  /note="cloning vector: pME18SFL3"

ORIGIN
Alignment Scores: 2.98e-96 Length: 2120
Pred. No.: 1457.50 Matches: 282
Score: 77.60% Conservative: 2
Percent Similarity: 77.05% Mismatch: 2
Best Local Similarity: 94.28% Indels: 80
Query Match: 9 Gaps: 1
DB:
US-10-054-313-1 (1-286) x AK096913 (1-2120)

```

```

Qy 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArg 20
Db 30 ATGAGCTGGCTTCTGTTCTTGGCCACAGAGTCGCTTGGCCGCTTGGCCCTTGGCCGCGC 89
Qy 21 GlySerArgGlyPheGlyMetPheTrpAlaValArgGlyArgGlyThrGlyValPhe 40
Db 90 GGCTCTCGGGTTCCGGATGTTCTATGCGTGGAGGGGCGGCAAGACCGGGTCTTT 149
Qy 41 LeuThrTrp----- 43
Db 150 CTGACCTGGTGAGTGGCCCGCCCGAGGGCGCTTCTCTCGCGCATCGGGTTCGAGGA 209
Qy 43 ----- 43
Db 210 GCGCGAGCGCGCTAACCCCGCGCGGAGCTCGCGGTCCGAGTCTTGTGCGCGCGC 269
Qy 43 ----- 43
Db 270 TGTGGCTGTGGGTGCTCGGGGACCGGCTCGTTTCTCTGCTCGCGCGGGGTTCAGG 329
Qy 43 ----- 43
Db 330 AGGCTCATGCTGGGACTGGAAACGACCGACCGAGACTTGTTTGAAATATCAGAAACGTCG 389
Qy 44 -----AsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLysPhe 61
Db 390 GTAAAGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCTCTGCTGCCAGATTTAAGAAATT 449
Qy 61 eAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSerGl 81
Db 450 TGCACACAGAGATGAGCGCTGGGCTTTGTCCAGGAATCTGCAAGCCGCGAAGTTTCAGA 509
Qy 81 uGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLysGly 101
Db 510 AGGCGATGAAATCAACATGACATGCAAGATCGGAGCGGAAAGCCAGGAGGACTCCGTGA 569
Qy 101 uProLeuAspGlyAspGlyHisGluSerAlaGlnProTrpAlaLysHisMetLysProSe 121
Db 570 GCCACTGGATGAGATGGACATGAAGCGCGAGCGCTATGCAAGCACATGAAGCCGAG 629
Qy 121 rValGluProAlaProValSerArgAspThrPheSerTrpMetGlyAspPheValVa 141
Db 630 CGTGAGCGCGCGCTCCAGTTAGCAGACACACGCTTTCTTACATGGAGACTTCGTCGT 689
Qy 141 lValTrpThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyVileGl 161
Db 690 GCTCTACACTGATGCTGCTCTGCTCAATGCGCGTAGAGCGCGGAGCAGGATCGG 749
Qy 161 yValTrpGlyProGlyHisProLeuAsnValGlyLysArgLeuProGlyArgGlnTh 181
Db 750 CGTTTACTGGGCGCGGCGCCATCTTTAAATGTAGGCATTAGACTTCTCGGCGCGCAGAC 809
Qy 181 rAsnGlnArgAlaGluIleHisAlaAlaCysValAlaIleGluGlnAlaLysThrGlnAs 201
Db 810 AAACCAAGAGCGGAAATTCATGACCTGCAAGCCCTGCAACAGCCAAAGAGACTCAAA 869
Qy 201 nIleAsnLysLeuValLeuTrpThrAspSerMetPheThrIleAsnGlyVileThrAsnTr 221
Db 870 CATCAATAAAGCTGTTCTGTATACACAGACAGTATGTTTACGATAAAATGTTAATACTG 929
Qy 221 pValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLys 241
Db 930 GGTTCAGGTTGGAAGAAAATGGGTGGAAGACAAGTCAGGAAAGAGGATGATCAACA 989
Qy 241 sGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisVa 261
Db 990 AGAGGACTTTGTGGCACTGGAGAGGCTTACCAGGGGATGGACATTCAGTGGATGCATGT 1049
Qy 261 lProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyVal 281
Db 1050 TCTGTGTCATTCGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGGAGC 1109
Qy 281 aLysGlnSerGluAsp 286

```

```

Db      1110 TARACATCGGAGAC 1125
|||||
RESULT 9
AF048993      1430 bp      mRNA      linear      ROD 21-NOV-1998
LOCUS      Mus musculus ribonuclease H1 (Rnh1) mRNA, complete cds.
ACCESSION      AF048993
VERSION      AF048993.1 GI:2935417
KEYWORDS
SOURCE      Mus musculus (house mouse)
ORGANISM
REFERENCE
AUTHORS      Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL      1 (bases 1 to 1430)
AUTHORS      Cerritelli, S.M., Fedoroff, O.Y., Reid, B.R. and Crouch, R.J.
TITLE      A common forty amino acid motif in eukaryotic RNases H1 and
JOURNAL      caulinovirus ORF VI proteins binds to duplex RNAs
AUTHORS      Nucleic Acids Res. (1998) In press
TITLE      2 (bases 1 to 1430)
JOURNAL      Cerritelli, S.M. and Crouch, R.J.
AUTHORS      Cloning, expression, and mapping of ribonucleases H of human and
TITLE      mouse related to bacterial RNase H1
JOURNAL      Genomics 53 (3), 300-307 (1998)
MEDLINE      99017966
PUBMED      9799596
REFERENCE      3 (bases 1 to 1430)
AUTHORS      Crouch, R.J. and Cerritelli, S.M.
TITLE      RNases H of lower eukaryotes: Saccharomyces cerevisiae,
JOURNAL      Schizosaccharomyces pombe, Neurospora crassa and Crithidia
AUTHORS      fasciculata
TITLE      (in) Crouch, R.J. and Toulme, J.-J. (Eds.);
JOURNAL      RIBONUCLEASES H;
REFERENCE      INSERM Editions, Paris, France (1998) In press
AUTHORS      4 (bases 1 to 1430)
TITLE      Cerritelli, S.M. and Crouch, R.J.
JOURNAL      Direct Submission
FEATURES
Location/Qualifiers
1..1430
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="dbEST:AA163387"
/db_xref="taxon:10090"
/chromosome="12"
/map="12A3"
/tissue_type="carcinoma"
/clone_lib="Stratagene 937317"
/dev_stage="embryonic"
1..1430
/gene="Rnh1"
1..49
/gene="Rnh1"
49..906
/gene="Rnh1"
/note="similar to bacterial RNases H1"
/codon_start=1
/product="ribonuclease H1"
/protein_id="AAC78562.1"
/db_xref="GI:2935418"
/translation="NRWLPLSRVTTLAVRLRGICGLGMFYVRRGRTGVFLSWS
ECKAVDPPARFKFKATEDEAWAFVSSSPDGSKGQSAHQKSOAKTSKRPREP
LGGSELPEFGPKHFRQTEPAVYSKDTFSYMGESVIYTDGCSNNGRKRABAGIG
VYWGPHLPUNGRQTNORAEIHAACKAIAKQAKNISKLVLYTDSMTFINGIT
NRVQGNKNGRTSTSGKDVINKEDFMELDTQGMNDIQNMHIPGHSFVGNEDRLA
REGAKQSD"
907..1430
/gene="Rnh1"
3'UTR
ORIGIN
Alignment Scores:
1409 bp      mRNA      linear      ROD 06-OCT-2003
BC019411
LOCUS      Mus musculus ribonuclease H1, mRNA (cDNA clone MGC:30220
DEFINITION

```

```

Pred. No.:      1,16e-77      Length:      1430
Score:      1197.50      Matches:      221
Percent Similarity:      86.71%      Conservative:      27
Best Local Similarity:      77.27%      Mismatches:      37
Query Match:      77.46%      Indels:      1
DB:      10      Gaps:      1
US-10-054-313-1 (1-286) x AF048993 (1-1430)
Qy      1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db      49 ATCCGCTGCTGCTGCCGCTGCCGACACTGCCGTCATCCCTGAGCGCA 108
Qy      21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40
Db      109 GGCATTTCGGGCTCGGGATGTTCTATCGGTGAGGAGGCGCAGACCGAGCTTC 168
Qy      41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
Db      169 CTGAGTTGGAGTGAGTGCACAAAGCCCAAGGTCGCGGTTTCTCTGCTGCCAGGTTTAA 228
Qy      61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db      229 TTTCACACAGAGATGAGCGCTGGGCTTTGTTCAGAGCTCTCAAGCCGGATGTTCA 288
Qy      81 GluGlyHisGluAsnGlnHisGlyGlnGlnSerGluAlaLysProGlyLysArgLeuArg 100
Db      289 AAAGGGCAGGAAAGTGCATGAGCAGAAAGTCCAGGCGAAGACCAAGCAGCGGCTCG 348
Qy      101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db      349 GAGCCCTG---GGTGAAGGGGAAGAACTTCAGAGCCAGGCGCAAGCAGCACACAG 405
Qy      121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
Db      406 GACACCGAGCCGCTGCTGTAGTGAGCAAGGACACATTTTCTATATGGAGAGTCAGTC 465
Qy      141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyLe 160
Db      466 ATTGTCTACACGATGGCTGTGCTCCAGTAATGGACGGAAGCGGCACAGCAGGAATT 525
Qy      161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyLeuArgLeuProGlyArgGln 180
Db      526 GCGGTTTACTGGGCGCCAGGCCACCCCTTAATGTAGGTATTAAGGCTTCTCTGGCGAC 585
Qy      181 ThrAsnGlnArgAlaGluLeuHisAlaAlaCysLysAlaLeuGluGlnAlaLysThrGln 200
Db      586 ACAAAACAGAGGCGCAGATGCCATGCGCTGCAAGGCCCATCATGCAAGCCAAAGGCTCAG 645
Qy      201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db      646 AACATCAGCAAGCTGGTTCGTACACAGACAGCATGTTTCCATCATATGGATACTAAC 705
Qy      221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240
Db      706 TGGGTTTCAGGCTCGAAGAGAATGCTGAGAGCAAGTACAGGAAAGATGTGATCAAC 765
Qy      241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Db      766 AAGAGGACTTCATGAGCTGAGAGCTGAGCTCCTCAGGCGCATGACATCCAGTGGATGCAC 825
Qy      261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280
Db      826 ATTCTCGTCACTCAGGATTTGTGGGCAATGAAGAGGCGCCACAGACAGATGGCAGCGA 885
Qy      281 AlalysGlnSerGluAsp 286
Db      886 CGCAAGCAGTCTGAGGAC 903

```

```

RESULT 10
BC019411
LOCUS      Mus musculus ribonuclease H1, mRNA (cDNA clone MGC:30220
DEFINITION

```

```

IMAGE:5133942), complete cds.
EC019411
VERSION 1
KEYWORDS GI:18044435
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.N., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,K., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schaeetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Lequellano,N.A., Peters,G.J.,
Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smallos,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
2238257
JOURNAL 12477932
MEDLINE 2
PUBMED 2 (bases 1 to 1409)
AUTHORS Strausberg,R.
DIRECT SUBMISSION
TITLE Submitted (19-DEC-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gcgabs@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulsegue, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LILNL at: http://image.llnl.gov
Series: IRAC Plate: 41 Row: n Column: 3
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 6755333.
FEATURES
source
location/Qualifiers
1..1409
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:30220 IMAGE:5133942"
/tissue_type="Liver, normal. 5 month old male mouse."
/clone_lib="NCI-CGAP_L19"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1..1409
/gene="Rnaseh1"
/db_xref="LocusID:19819"

```

/db\_xref="MGI:1335073"  
 17..874  
 /codon\_start=1  
 /product="ribonuclease H1"  
 /protein\_id="AAH19411.1"  
 /db\_xref="GI:18044436"  
 /db\_xref="LocusID:19819"  
 /translation="MRWLLPLSLSTVTLAVVLRRLGCGIGMFYAVRGRRTGTVFLSWS  
 ECAQVDPRPAARFKKFKFATEDAWAFVRSRSPDGSGSHEQKSVKRSRREP  
 LGSEBLPEPGPKHTQDTEPAAVSKDTFSYMGESVIVYTDCCSSNGKRRAGIG  
 VYNGPGLNVGIRLPGROTNPAETHAACKAIMOAKQNIKSLVLYTDSMFTINGIT  
 NWQCKKNGKWRSTSTGDKVINKDFMELDELQGMQIOWHPIHSGSFVGNBEADRLA  
 REKAQSD"  
 89..232  
 /note="COG3341; Region: Predicted double-stranded  
 RNA/RNA-DNA hybrid binding protein [General function  
 prediction only]"  
 /db\_xref="CDD:COG3341"  
 428..859  
 /note="RnH1; Region: Ribonuclease H1 [DNA replication,  
 recombination, and repair]"  
 /db\_xref="CDD:COG0328"

Alignment Scores:  
 Pred. No.: 2,22e-77 Length: 1409  
 Score: 1193.50 Matches: 220  
 Percent Similarity: 86.36% Conservative: 27  
 Best Local Similarity: 76.92% Mismatches: 38  
 Query Match: 77.20% Indels: 1  
 DB: 10 Gaps: 1

US-10-054-313-1 (1-286) x BC019411 (1-1409)  
 QY 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuAlaLeuProCysArgArg 20  
 Db 17 ATGGCTGGCTGCTGCGCTGCTCCGCACAGTGCACACTGCGCGTCGACCGCTGAGCGCA 76  
 QY 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgGlyThrGlyValPhe 40  
 Db 77 GCGATTTCGGGGCTCGGGAGTGTCTATGCGGTGAGGAGCGCCGACGAGCGAGTCTTC 136  
 QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60  
 Db 137 CTGAGTTGGAGTGAGTGCAGAGCCAGTGCACCGGTTTCTGCTGTCAGGTTTAAAGAA 196  
 QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80  
 Db 197 TTGGCCACAGAGATGAGCGCTGGCCCTTTGCTAGGAGCTCTTCAAGCCCGGATGGTCA 256  
 QY 81 GluGlyHisGluAsnGlnHisGlyGlnGlnSerGluAlaLysProGlyLysArgLeuArg 100  
 Db 257 AAAGGGCAGGAAGATGTCACATGACAGAGTGCACAGTGAAGACACAGCAGCGCGCTCGG 316  
 QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120  
 Db 317 GAGCCCTCTG---GGTGAAGGGGGAAGAACTTCCAGCGCCGCGGCCCAAGCACACAGCAG 373  
 QY 121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140  
 Db 374 GACACCGAGCCCGCTGCTGTAGTGAGCAAGCACACGTTTCTTATATGGGAGAGTCAGTC 433  
 QY 141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgLysProArgAlaGlyTle 160  
 Db 434 ATTGTCTACCGGATGGCTGTGTCTCCAGTAATGAGCAAGGAGCGGCGCAGCAGGAAT 493  
 QY 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180  
 Db 494 GCGTTTACTGGGCGCCAGGCCACCCCTTAATAGTAGTAAGAGCTTCTCTGGCGCAG 553  
 QY 181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200  
 Db 554 ACAAAACAGAGGGCGGAGATCCATGTCAGCTGCAGGCGCATCATGCAAGCCAGGCTCAG 613

QY	201	AsnIlAasNlySleuValLeuThyTrAspSerMetPheThrIleAsnGlyIleThrAsn	220
DB	614	AACATCAGCAAGCTGGTGTCTGTGCACACAGCATGTTCCACCATCAATGGATAACTAAC	673
QY	221	TTPVaGlncNGlyTriPLysLysAsnGlyTriPLysThrSerAlaGlyLysGluValIleAsn	240
DB	674	TGGGTCAGGGCTCGAAGAAGATGGCTGGAGACAAGTAGTACAGGAAGAAGATGATCAAC	733
QY	241	LysGluAspPheValIlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis	260
DB	734	AAGGAGACTTCACTGGAGCTGGACAGACTCACTCAGGCGCATGGACATCCAAATGGATGCAC	793
QY	261	ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly	280
DB	794	ATTCTTGCTCACTCAGGATTTGTGGCAATAGAGAGCCGACAGACTGGCAGCGGAAGGA	853
QY	281	AlaLysGlnSerGluAsp	286
DB	854	GCGAAGCAGTCTGAGGAC	871
RESULT	11		
LOCUS	AC098850	178965 bp DNA linear PRI 30-DEC-2001	
DEFINITION	Homo sapiens chromosome 17, clone RP11-92B11, complete sequence.		
ACCESSION	AC098850		
VERSION	AC098850.2 GI:17646937		
KEYWORDS	HTG.		
SOURCE	Homo sapiens (human)		
ORGANISM			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
REFERENCE			
AUTHORS	Barren,B., Linton,L., Nusbaum,C. and Lander,E.		
TITLE	Homo sapiens chromosome 17, clone RP11-92B11		
JOURNAL	Unpublished		
REFERENCE			
AUTHORS	Barren,B., Linton,L. (to 178965)		
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chararo,B., Choepel,Y., Colangelo,M., Collins,S., Collimore,A., Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Govette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazeres,R., Landers,T., Lehoczyk,J., Levine,R., Liu,G., Maclean,C., MacDonald,P., Major,O., Marquis,N., Matthews,C., McCarthy,M., Mcwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Menga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnapack,R., Seaman,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zemlek,L., Zimmer.A. and Zody,M.			
Direct Submission			
TITLE			
JOURNAL			
Submitted (03-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA			
3 (Bases 1 to 178965)			
REFERENCE			
AUTHORS	Barren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chararo,B., Choepel,Y., Colangelo,M., Collins,S., Collimore,A., Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Govette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazeres,R., Landers,T., Lehoczyk,J., Levine,R., Liu,G., Maclean,C., MacDonald,P., Major,O., Marquis,N., Matthews,C., McCarthy,M., Mcwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Menga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnapack,R., Seaman,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zemlek,L., Zimmer.A. and Zody,M.		

```
repeat_region 7310..7435 /rpt_family="LIME"
repeat_region 7527..7703 /rpt_family="LIME"
repeat_region 7704..7993 /rpt_family="AluJo"
repeat_region 7994..8309 /rpt_family="LIME"
repeat_region 8319..8807 /rpt_family="L2"
repeat_region 8893..9579 /rpt_family="L2"
repeat_region 9341..9994 /rpt_family="GA-rich"
repeat_region complement(10332..10491) /rpt_family="MER3"
repeat_region complement(11158..11392) /rpt_family="L3"
repeat_region 11635..12721 /rpt_family="L1MC5"
repeat_region 12814..13108 /rpt_family="AluY"
repeat_region complement(13131..13216) /rpt_family="L1P"
repeat_region 13203..13257 /rpt_family="L1PA4"
repeat_region 13255..15125 /rpt_family="L1PA4"
repeat_region complement(15190..15383) /rpt_family="MIR"
repeat_region 15752..16394 /rpt_family="L1MB7"
repeat_region 16392..16515 /rpt_family="L1MB7"
repeat_region 16513..17530 /rpt_family="L1MB7"
repeat_region complement(17531..17840) /rpt_family="AluSx"
repeat_region 17841..19946 /rpt_family="L1MB7"
repeat_region 20113..20529 /rpt_family="MSTB1"
repeat_region 20934..21416 /rpt_family="LTR65"
repeat_region 21417..21594 /rpt_family="AluSx"
repeat_region 21929..22229 /rpt_family="AluSx"
repeat_region complement(22689..23156) /rpt_family="LTR37B"
repeat_region complement(24183..24502) /rpt_family="AluSx"
repeat_region complement(26205..26386) /rpt_family="MIR"
repeat_region 26671..26697 /rpt_family="(TAA)n"
repeat_region 27404..27705 /rpt_family="AluSx"
repeat_region complement(28018..28146) /rpt_family="MER5A"
repeat_region 28157..28214 /rpt_family="AT_rich"
repeat_region 28714..28743 /rpt_family="AT_rich"
repeat_region 28825..28870 /rpt_family="AT_rich"
repeat_region 31073..31292 /rpt_family="L1ME4A"
repeat_region 31553..31858 /rpt_family="L3"
repeat_region 32320..32351 /rpt_family="AT_rich"
repeat_region 32720..33020 /rpt_family="AT_rich"

/rpt_family="AluJo"
33053..33370
/rpt_family="AluYb8"
complement(33693..34052)

Alignment Scores:
Pred. No.: 1.24e-73 Length: 178965
Score: 1174.00 Matches: 236
Percent Similarity: 88.36% Conservative: 7
Best Local Similarity: 85.82% Mismatches: 18
Query Match: 75.94% Indels: 16
DB: 9 Gaps: 3

US-10-054-313-1 (1-286) x AC098850 (1-178965)
Qy 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db 59696 ATGAGCGCGCTTCTGCTCCCTGGCGGC-AGAGTCGCTTGGCCGCTTGGCCGCGC 59754
Qy 21 GlySerArgGlyPheGlyMetPheTrpAlaValArgGlyArgGlyThrGlyValPhe 40
Db 59755 ---TCTCGCGGGTTCGGGATGTTCTATGCGTGAGAGGGCGGAGTCCGCGTCTTT 59811
Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLys 60
Db 59812 CTGACCGGGAATGAGTGCAGGACAGGTCGACCGGTTTCTGCTGCAGATTAAAGAG 59871
Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 59872 TTTGCCACAGGATGAGGCTGGGACTTTGTTCAGAAATCTCGACGCCCGAAGTTTCA 59931
Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db 59932 GRAGGGCAGGAAATCAACATGACAGGAATCGAGACGAGGAGCCAGGACGCTCGT 59991
Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTrpAlaLysHisMet---- 118
Db 59992 GAGCCACTGGATGGAGATGAGATGAAGCGCAGAGCCATATGC-AAGCACATGAAGCCG 60050
Qy 119 -----LysProSerValGluProAlaProAlaProValSerArgAspThr 132
Db 60051 AGCGTGAAGCGCGGTGAGCGCGGTGAGCGCGGCTCCAGTTAGCAGACACG 60110
Qy 133 PheSerTrpMetGlyAspPheValValValValValValValValValValValVal 152
Db 60111 TTTTCTACATGGGAGACTTCGTCGCTGCTACGCTGATGGCTGCTCCAGTAATGGG 60170
Qy 153 ArgArgLysProArgAlaGlyIleGlyValValValValValValValValValVal 172
Db 60171 CGTAGAAGCCACGAGCAGGAATCCGCTTTACTGGGGCGGGGCTATCTTTAACTGTA 60230
Qy 173 GlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGluIleHisAlaLysLys 192
Db 60231 GGCATTAGACTTCCTGGGTGGCAGACAAACCAAGAGCGGAAATTCATGAGCGCTGCAAG 60290
Qy 193 AlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuValValValVal 212
Db 60291 GCCATTGAACAAAGCAGCAGACTCAGAAATCAATAAATCTGTTCTGTATACAGCAGATG 60350
Qy 213 PheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLysAsnGlyTrpLysThr 232
Db 60351 TTTACGATAAATGGTATAACTAACTGGGTTTCAAGTTTGAAGAAATTCGTTGGAAGACA 60410
Qy 233 SerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGln 252
Db 60411 AGTCAGGGAAGAGGTGATCAACAAAGAGGACTTCGTGACCTGGAGAGCTGACCCAG 60470
Qy 253 GlyMetAspIleGlnTrp-----MethHisValPro 262
Db 60471 GGGATGGACATTCACTGGCGGAGTATCTTAAATGTTTCATGTCCA 60515

RESULT 12
AC022596/c
```

```

LOCUS          AC022596          192337 bp      DNA      linear      PRI 23-SEP-2000
DEFINITION     Homo sapiens chromosome 17, clone RP11-219A15, complete sequence.
ACCESSION      AC022596
VERSION        AC022596.9  GI:10280853
KEYWORDS       HTG.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 192337)
AUTHORS        Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
TITLE          Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,
JOURNAL        Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
AUTHORS        Chospel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
TITLE          DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
JOURNAL        Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
AUTHORS        Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
TITLE          Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
JOURNAL        Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
AUTHORS        Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
TITLE          McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
JOURNAL        Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
AUTHORS        Pierre,N., Pisaní,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
TITLE          Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
JOURNAL        Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
AUTHORS        Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
TITLE          Zimmer,A. and Zody,M.
JOURNAL        Direct Submission
REFERENCE      Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
AUTHORS        Research, 320 Charles Street, Cambridge, MA 02141, USA
               3 (bases 1 to 192337)
               Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
               Anderson,S., Barna,N., Bastien,V., Bada,F., Boguslavskiy,L.,
               Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
               Chospel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
               DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
               Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
               Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
               Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
               Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G.,
               Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
               McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
               Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
               O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K.,
               Pierre,N., Pisaní,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
               Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
               Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
               Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
               Tirrell,A., Travers,M., Triggilio,J., Vassiliev,H., Viel,R., Vo,A.,
               Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
               Zimmer,A. and Zody,M.
               Direct Submission
TITLE          Submitted (23-SEP-2000) Whitehead Institute/MIT Center for Genome
JOURNAL        Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT        On Sep 23, 2000 this sequence version replaced gi:9795639.
               All repeats were identified using RepeatMasker:
               Smit, A.F.A. & Green, P. (1996-1997)
               http://ftp.genome.washington.edu/RM/RepeatMasker.html
               ----- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIGR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
               ----- Project Information
               Center project name: L4191
               Center clone name: 219_A_15
               -----
               Location/Qualifiers
               1..192337

```

```

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-219A15"
/clone_lib="RPCI-11 Human Male BAC"
568..788
/rpt_family="MLT1C"
1804..1957
/rpt_family="FRAM"
1964..2014
/rpt_family="MIR"
complement(2029..2319)
/rpt_family="LTR16C"
2463..2535
/rpt_family="MIR"
complement(2614..2700)
/rpt_family="MIR3"
complement(2938..3185)
/rpt_family="L1MC/D"
complement(4802..4869)
/rpt_family="L2"
complement(4871..5058)
/rpt_family="LTR16B"
complement(5150..5227)
/rpt_family="LTR16A"
5353..5597
/rpt_family="AluJo"
5697..5863
/rpt_family="MLT1J"
complement(5909..6210)
/rpt_family="AlusX"
complement(6235..6609)
/rpt_family="L1MB3"
6610..6687
/rpt_family="AlusG/X"
complement(6688..6802)
/rpt_family="L1MB3"
7399..7751
/rpt_family="MSTD"
7977..8102
/rpt_family="MER5A"
complement(8265..8596)
/rpt_family="L1MB3"
complement(10173..10360)
/rpt_family="MIR"
complement(10370..10452)
/rpt_family="MIR"
10648..11010
/rpt_family="L1MB4"
complement(11168..11273)
/rpt_family="L1MC4"
complement(11274..11583)
/rpt_family="AlusX"
complement(11584..11603)
/rpt_family="L1MC4"
12175..12289
/rpt_family="MIR"
complement(12402..12464)
/rpt_family="L2"
13177..13351
/rpt_family="MER69"
13950..14083
/rpt_family="MIR"
complement(14660..15282)
/rpt_family="L1MC4a"
complement(15834..15968)
/rpt_family="MIR3"
complement(16176..16237)
/rpt_family="L2"
17027..17098
/rpt_family="MIR"

```

```

repeat_region complement(17196..17328)
/rpt_family="L2"
repeat_region complement(17390..17492)
/rpt_family="L2"
repeat_region 18087..18429
/rpt_family="L2"
repeat_region 18524..18675
/rpt_family="MER3"
repeat_region 18848..19158
/rpt_family="AluSp"
repeat_region complement(19554..19800)
/rpt_family="AluSx"
repeat_region complement(20989..21284)
/rpt_family="AluY"
repeat_region 21286..21365
/rpt_family="MER3"
repeat_region 21393..21576
/rpt_family="L2"
repeat_region 21627..21701
/rpt_family="MIR"
repeat_region 21770..22001
/rpt_family="LTR14A"
repeat_region complement(22014..22401)
/rpt_family="L1MA4A"
repeat_region 23317..23339
/rpt_family="AT rich"
repeat_region complement(23367..23502)
/rpt_family="MIR"
repeat_region 24118..24141
/rpt_family="(TCTG)n"
repeat_region complement(24473..24616)
/rpt_family="MIR3"
repeat_region complement(25652..26053)
/rpt_family="L2"
repeat_region 28293..28376
/rpt_family="MIR3"
repeat_region 34302..34408
/rpt_family="MIR"
repeat_region complement(35693..35781)
/rpt_family="MIR"
repeat_region 37220..37405
/rpt_family="L2"
repeat_region complement(37431..37554)
/rpt_family="MIR"
repeat_region 38931..39908
/rpt_family="LTR5"
repeat_region 40343..40448
/rpt_family="MIR"
repeat_region 40552..40629
/rpt_family="MIR"
repeat_region complement(41149..41275)
/rpt_family="MIR"
repeat_region complement(41355..41506)
/rpt_family="MIR3"

```

## Alignment Scores:

```

Pred. No.: 1.34e-73 Length: 192337
Score: 1174.00 Matches: 236
Percent Similarity: 88.36% Conservative: 7
Best Local Similarity: 85.82% Mismatches: 18
Query Match: 75.94% Indels: 16
DB: 9 Gaps: 3

```

US-10-054-313-1 (1-286) x AC022596 (1-192337)

```

Qy 1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db 191561 ATGAGCGGGTTCGTCCCTGGGGC-AGAGTCGCTTGGCCCTTCCTCCCGCCG 191503
Qy 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyArgGlyThrGlyValPhe 40
Db 191502 ---TCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGCGGCAAGTCCCGGGTCTTT 191446

```

```

Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLys 60
Db 191445 CTGACCGGGAATGAGTGCACAAAGCACAGGTGCACCGGTTCTGCTGCAGATTTAAGAAG 191386
Qy 61 PheAlaThrGluAspGluAlaTTPAlaPheValArgLysSerAlaSerProGluValSer 80
Db 191385 TTTCACACAGAGGATGAGGCTGGGACTTTGTACGAAATCTGCAAGCCCGAAGTTTCA 191326
Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db 191325 GAAGGCGAGGAAATCAACATGGAAGAATCGGAGAGCAAGAACCCAGCAGCAGCTCCGT 191266
Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMet----- 118
Db 191265 GAGCCACTGGATGGAGATGGAAGCGCAGCATATGC-AAGCACATGAAGCCG 191207
Qy 119 -----LysProSerValGluProAlaProValSerArgAspThr 132
Db 191206 AGCGTGAAGCCGAGCGGTGAGCGCGGTGAAGCGCGCGCTCAGTTAGCAGAGACAG 191147
Qy 133 PheSerTyrMetGlyAspPheValValTyrThrAspGlyCysSerSerSerAsnGly 152
Db 191146 TTTTCTACATGGGAGACTTCGTCGTCGTCTACGCTGATGCTGCTCTCCAGTAATGGG 191087
Qy 153 ArgArgLysProArgAlaGlyLysGlyValTyrTrpGlyProGlyHisProLeuAsnVal 172
Db 191086 CGTAGAAGCCACGAGCAGGAAATCGCGTTTACTGGGGCGCGGCTATCCTTTAATGTA 191027
Qy 173 GlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGluLysHisAlaAlaCysLys 192
Db 191026 GGCATTAGACTTCTCTGGGTGGCAGCAAAACCAAGAGCGGAAATTCATGAGCCTGCAAG 190967
Qy 193 AlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMet 212
Db 190966 CCCATTGAACAAGCAGCAGACTCAGAAATCAATAAATGTTCTGTTATACAGACGATG 190907
Qy 213 PheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLysAsnGlyTrpLysThr 232
Db 190906 TTTACGATAAATCGTATAAATACTAACTGGGTTCAAGTTTGAAGAAAAATGGTGAAGACA 190847
Qy 233 SerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGln 252
Db 190846 AGTCAGGGAAGAGGATGATCAACAAGAGAGACTTCGTGGCAGCTGGAGAGCTGCCCG 190787
Qy 253 GlyMetAspIleGlnTrp-----MethHisValPro 262
Db 190786 GGGATGGACATTCAGTGGCGAGTATCTTAAATGTTTCATGTCCTCA 190742
RESULT 13
AC126352/6
LOCUS 211403 bp DNA linear PRI 19-MAR-2003
DEFINITION Homo sapiens chromosome 17, clone CTD-2509K7, complete sequence.
ACCESSION AC126352
VERSION AC126352.6 GI:29124211
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 211403)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone CTD-2509K7
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 211403)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Bouckgater,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeAtellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A.,
Karatas,A., Kellis,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,

```













Qy	21	GlySerArgGlyPheGlyMetPheTyrAlaValArgArgIyArgIyThrGlyValPhe	40
Db	77541	TGCTCTCAACGGGTTCGGGATGTTCTATGCGGTGAGAGGGCGGCAAGACCGGGTCTTT	77600
Qy	41	LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheIysIys	60
Db	77601	CTGACCGGGAATGAGTGCAAAGCACACAGAGTGGACCGGTTCTTGCTGCCAGATTAAAGNAG	77660
Qy	61	PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer	80
Db	77661	TTTGCCACACAGAGATGAGACCTGGGACTTTGTTCAGAAATCTTCGAAGCCCGGAAGTTTCA	77720
Qy	81	GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGluLysArgGluArg	100
Db	77721	GAAGGCAGAGAAATCAACATGGACACAGAAATCAGACACAAAGCCAGCAGCTCCGT	77780
Qy	101	GluProIeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetIysPro	120
Db	77781	GAGCCACTGATGGATGGAGATGAAGCGAGAGCCATATGC-AAGCACATGAAGCCG	77839
Qy	121	SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal	140
Db	77840	AGGGTAGACCGCGCGCTCCAGTTAGCAGAGACAGAGTTTCTCTACATGGGAGACTTCGTC	77899
Qy	141	ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle	160
Db	77900	GTCTGCTACGCTGATGGCTGCTGCTCCAGTAAATGGCGGTAGAAGGCCACGACGAGCAATC	77959
Qy	161	GlyValTyrTrpGlyProGlyIysProIeuAsnValGlyIleArgLeuProGlyArgGln	180
Db	77960	CGCGTTTACTGGGGCGGGCGCTATCCTTTAAATGTAGGCATTAGACTTCTCTGGTGGCAG	78019
Qy	181	ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln	200
Db	78020	ACAAACAAAGACGGGAAATTCATGCAGCGCTACAAGGCCANTGAAACAAGCAAGACTCAA	78079
Qy	201	AsnIleAsnLysIeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn	220
Db	78080	AACATCAATAAATCGTTCTGTATACAGACAGATGTTTACGATAAATGGTGT-----AAC	78135
Qy	221	TrpValGlnGlyTrpLysLysAsnGlyTyrIlyThrSerAlaGlyLysGluValIleAsn	240
Db	78136	TGGGTTTCAGGTTGGGAAGAAAAATTCGATGGAAAGACAAGTGCAGGGGAAAGAGGTGACCAAC	78195
Qy	241	LysGluAspPheValAlaIleuGluArgLeuThrGlnGlyMetAspIleGlnTrp-----	258
Db	78196	AAAGAGACTTCGTGGCACTGGAGAGGCTGACCCAGGGGATGGACATTCAGTGGGCGAGT	78255
Qy	259	-----MethHisValPro 262	
Db	78256	ATCTTAAATGTTTCATGTCCCA 78276	

Search completed: March 22, 2004, 23:32:51  
Job time : 3019 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 22, 2004, 21:36:56 ; Search time 330 Seconds

(without alignments)  
3681.772 Million cell updates/sec

Title: US-10-054-313-1

Perfect score: 1546

Sequence: 1 MSWLLFLAHLVALAALPCR.....FIGNEADRLAREGAKQSD 286

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cpn2\_1/USPTO\_spool/US10054313/runat\_22032004\_140415\_28060/app\_query.fasta\_1.455  
-DB=N\_Geneseq\_29Jan04 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=tblowsum2 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10054313 @CGN.1.1.352 @runat\_22032004\_140415\_28060 -NCFU=6 -ICFU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_29Jan04:\*

- 1: Geneseqn1980s:\*
- 2: Geneseqn1990s:\*
- 3: Geneseqn2000s:\*
- 4: Geneseqn2001as:\*
- 5: Geneseqn2001bs:\*
- 6: Geneseqn2002as:\*
- 7: Geneseqn2003as:\*
- 8: Geneseqn2003bs:\*
- 9: Geneseqn2003cs:\*
- 10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1526	98.7	2129	5	ABV30263 Human pro
2	1523	98.5	1150	3	Aaz51265 Human RNA
3	1522	98.4	965	7	ADA05725 Human NOV
4	1218	78.8	979	9	ADC39107 Novel hum
5	1045.5	67.6	631	7	ADA05719 Human NOV
6	1045.5	67.6	716	7	ADA05717 Human NOV
7	979.5	63.4	586	7	ADA05721 Human NOV
8	905	58.5	545	5	ABV45787 Human pro

9	792	51.2	457	7	ADA05723
10	666.5	43.1	691	5	ABV15988
11	606	39.2	519	9	ADD34573 Mouse mit
12	571	36.9	882	9	ADD34572 Mouse mit
13	508.5	32.9	389	3	AAH30241 Human col
14	468	30.3	436	8	ACH42294 Human col
15	434	28.1	473	7	ABX41815 Bovine ES
16	422	27.3	1071	4	ABL03247 Drosophi
17	399	25.8	407	7	ABX38735 Bovine ES
18	364	23.5	764	6	ABX77198 Frog embr
19	356.5	23.1	3290	4	ABL03246 Drosophi
20	272	17.6	119211	4	Aaf28553 Genomic f
21	249.5	16.1	1047	7	ACC61723 Gene sequ
22	244	15.8	378	2	AAT24248 Human gen
23	239	15.5	7874	2	AAT20568 Polynucle
24	236.5	15.3	462	5	Aaf94380 Haemophil
25	236.5	15.3	462	6	ABK64944 DNA encod
26	236.5	15.3	1001	4	Aaf91449 Haemophil
27	236.5	15.3	1001	6	ABK37829 DNA sequ
28	236.5	15.3	110000	2	ABV42063_01
29	235.5	15.2	2048	2	AAX98244 Contig 13
30	234.5	15.2	435	7	ABZ40958 N. gonorr
31	229.5	14.8	36471	3	AAa81453 N. mening
32	229.5	14.8	110000	3	AAa81489_2
33	229.5	14.8	349980	3	AAF21611 Neisseria
34	221.5	14.3	801	2	AAX98246 Nucleotid
35	217	14.0	1509	8	ADA30478 DNA encod
36	213.5	13.8	465	2	AAQ96268 E.coli ri
37	211.5	13.7	465	2	AAQ96267 E.coli ri
38	209.5	13.6	757	2	AAQ12780 Encodes s
39	203	13.1	349980	6	ABQ81849 Bifidobac
40	200.5	13.0	471	7	ACF69459 Photorhab
41	200.5	13.0	110000	7	ACF67367_24
42	200.5	13.0	110000	7	ACF65386_4
43	190.5	12.3	630	8	ADBI11347_4
44	190.5	12.3	630	8	ADBI11349
45	190.5	12.3	110000	8	ADBI2064_15

# ALIGNMENTS

RESULT 1	ABV30263
ID	ABV30263 standard; cDNA; 2129 BP.
XX	
AC	ABV30263;
XX	
DT	16-SEP-2002 (first entry)
XX	
DE	Human prostate expression marker cDNA 30254.
XX	
KW	Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW	pharmacogenomic marker; gene; ss.
XX	
OS	Homo sapiens.
XX	
PN	W0200160860-A2.
XX	
PD	23-AUG-2001.
XX	
PF	20-FEB-2001; 2001WO-US005171.
XX	
PR	17-FEB-2000; 2000US-0183319P.
PR	16-MAR-2000; 2000US-0189862P.
PR	23-MAY-2000; 2000US-0207454P.
PR	09-JUN-2000; 2000US-0211314P.
PR	18-JUL-2000; 2000US-0219007P.
PR	13-DEC-2000; 2000US-0255281P.
PA	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX	
FI	Schlegel R, Endege WO, Monahan JE;
XX	

DR WPI; 2001-662795/76.  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer.  
XX  
XX  
PS Claim 1; Page 6567-6568; 11750pp; English.  
CC The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for: (a) assessing whether  
CC a patient is afflicted with prostate cancer; (b) monitoring the  
CC progression of prostate cancer in a patient; (c) assessing the efficacy  
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing  
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
CC determining whether prostate cancer has metastasized in a patient; (h)  
CC assessing the aggressiveness or indolence of prostate cancer in a patient  
CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker  
XX  
SQ Sequence 2129 BP; 609 A; 510 C; 578 G; 427 T; 0 U; 5 Other;  
  
Alignment Scores:  
Pred. No.: 1.5e-125 Length: 2129  
Score: 1526.00 Matches: 282  
Percent Similarity: 99.30% Conservative: 2  
Best Local Similarity: 98.60% Mismatches: 2  
Query Match: 98.71% Indels: 0  
DB: 5 Gaps: 0  
  
US-10-054-313-1 (1-286) X ABV30263 (1-2129)  
QY 1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArg 20  
DB 112 ATGAGCTGCTTCTGTTCTGCGCCACAGATCGCTTGGCGCTTGGCGCGCGC 171  
QY 21 GlySerArgGlyPheGlyWetPheTrpAlaValArgArgGlyArgGlyThrGlyValPhe 40  
DB 172 GGCCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGAGGGCGCGCAAGACCGGGGTCTTT 231  
QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60  
DB 232 CTGACCTGGAATGAGTGCAGACACAGATGGACCGTTCCTGCTGCCAGATTTAAGAAG 291  
QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80  
DB 292 TTTGCCACAGAGATGAGCCTGGCGCTTTGTGAGGAAATCTGCAAGCCCGGAACTTCA 351  
QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100  
DB 352 GAAGGCGATGAAATCAATGACATGACAAAGATCGAGGCGAAGCCAGCAGCATTCGT 411  
QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTrpAlaLysHisMetLysPro 120  
DB 412 GAGCCACTGGATGGATGACATGAAAGCGCAGACCGCTATGCAAGACACATGAAGCGC 471  
QY 121 SerValGluProAlaProValSerArgAspThrPheSerTrpMetGlyAspPheVal 140  
DB 472 AGCGTGAGCGCGCGCTCAGTTAGCAGACACAGCTTTTCTTACATGGAGACATTCGTC 531  
QY 141 ValValTrpThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyLe 160  
DB 532 GTGCTCTACACTGATGGCTGCTCTCCAGTAATGGCGTAGAAGCGCGCGAGCAAGATC 591  
QY 161 GlyValTrpGlyProGlyHisProLeuAsnValGlyLysArgLeuProGlyArgGln 180  
DB 592 GGCCTTTACTGGGGCGAGCCCATCTTTAAATGTAGGCATTAGACCTTCTGGCGCGGAG 651  
QY 181 ThrAsnGlnArgAlaGluLeuHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200  
DB 652 ACAACCAAGAGCGGAATTCATGAGCTGCAAGCCATTGCAACCAAGACTCAA 711  
  
QY 201 AsnIleAsnLysLeuValLeuTrpThrAspSerMetPheThrIleAsnGlyIleThrAsn 220  
DB 712 AACATCAATAAATCTGTTCTGTATACAGACAGTAATGTTTACGATAAATGGTAACTAAC 771  
QY 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240  
DB 772 TGGGTTCAAGGTTGGAAGAAAATGGTGAAGACAAAGTCCAGGGAAGAGGTTGATCAAC 831  
QY 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260  
DB 832 AAGAGAGACTTTTGGCAGCTGGAGAGCTTACCCAGGGATGGACATTCATGGTGGATGCAT 891  
QY 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280  
DB 892 GTTCTCTGCTCATTCGGGATTTATAGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGGA 951  
QY 281 AlaLysGlnSerGluAsp 286  
DB 952 GCTAAACAAATCGGAAGAC 969  
  
RESULT 2  
AAZ51265  
ID AAZ51265 standard; cDNA; 1150 BP.  
XX  
AC AAZ51265;  
XX  
DT 06-JUN-2000 (first entry)  
XX  
DE Human RNA-associated protein-16 (RNAAP-16) encoding cDNA.  
XX  
KW RNA-associated protein; RNAAP; human; clone 2073417; cytosolic;  
KW immunosuppressive; antiinflammatory; hepatocytic; neuroprotective;  
KW antiarteriosclerotic; antiproliferative; antiproliferative; anti-HIV;  
KW antiallergic; antirheumatic; antithrombotic; antithrombotic; autoimmune;  
KW antimicrobial; cell proliferative disorder; inflammation; cirrhosis;  
KW actinic keratosis; bursitis; arteriosclerosis; artherosclerosis;  
KW hepatitis; myelofibrosis; primary thrombocythemia; psoriasis; cancer;  
KW mixed connective tissue disease; MCTD; HIV; uveitis; Crohn's disease;  
KW allergy; rheumatoid arthritis; parasitic infection; ss.  
XX  
OS Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH CDS 85..945  
FT /\*tag= a  
FT /product= "Human RNA-associated protein-16"  
FT /note= "Derived from ISLNT01 library"  
FT sig\_peptide 85..138  
FT /\*tag= b  
FT mat\_peptide 139..942  
FT /\*tag= c  
FT /product= "Mature human RNA-associated protein-16"  
XX  
XX WO200011171-A2.  
XX  
XX PD 02-MAR-2000.  
XX  
XX PF 20-AUG-1999; 99WO-US019361.  
XX  
XX PR 21-AUG-1998; 98US-0097550P.  
XX  
XX PR 12-JAN-1999; 99US-0115639P.  
XX  
XX PA (INCY-) INCYTE PHARM INC.  
XX  
XX PI Hillman JL, Yue H, Tang YT, Corley NC, Guegler KJ, Gorgone GA;  
XX Patterson C, Baughn WR, Lal P, Bandman O, Reddy R, Azimzai Y;  
XX PI Shih LL, Yang J, Lu DAM;  
XX  
XX DR WPI; 2000-237651/20.  
XX  
XX DR P-PSDB; RAY70235.  
XX  
XX PT Human RNA-associated proteins useful in diagnosing, treating and  
XX preventing cell proliferative, autoimmune, inflammatory and infectious

PT disorders.

PS Claim 9; Page 117-118; 123pp; English.

XX The present sequence is the cDNA encoding human RNA-associated protein-16 (RNAP-16), identified in Incyte clone 2073417, derived from iSLTNOT01 library. It is expressed in nervous, reproductive, gastrointestinal, cardiovascular and haematopoietic/immune tissues. It has cytosolic, immunosuppressive, antiinflammatory, antiarteriosclerotic, hepatotropic, keratolytic, neuroprotective, antipsoriatic, anti-HIV, antiallergic, antirheumatic, virucide, antiarthritic, ophthalmological and antimicrobial activity. RNAP antibodies are useful for diagnosis of diseases associated with altered expression or activity of RNAP. It is used to treat cell proliferative, autoimmune, inflammatory and infectious disorders, like actinic keratosis, bursitis, arteriosclerosis, artherosclerosis, cirrhosis, hepatitis, myelofibrosis, mixed connective tissue disease (MCTD), psoriasis, primary thrombocythemia and cancer, HIV, allergies, rheumatoid arthritis, uveitis, Crohn's disease, and bacterial, viral and parasitic infections

SQ Sequence 1150 BP; 302 A; 250 C; 352 G; 246 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 1.3e-125 Length: 1150  
Score: 1523.00 Matches: 281  
Percent Similarity: 99.30% Conservative: 3  
Best Local Similarity: 98.25% Mismatches: 2  
Query Match: 98.51% Indels: 0  
DB: 3 Gaps: 0

US-10-054-313-1 (1-286) x AA251265 (1-1150)

Qy 1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20  
Db 85 ATGAGCTGGCTTCTGTTCTTCTGCCCCACAGTCGGCTTGGCCGCTTGCCTTGGCCGCG 144  
Qy 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyArgLysThrGlyValPhe 40  
Db 145 GGCTCTCGCGGGTTCGGGATGTTCTATGCGGTGAGGAGGGCGCGACAGCCGGGCTCTTT 204  
Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAlaPheArgGlyPheProAlaAlaArgPheLys 60  
Db 205 CTGACCTGGAATGAGTGACGACAGACAGGTGGACCGGTTCTCTGCTGCCAGATTTAAGAAG 264  
Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80  
Db 265 TTGGCCACAGAGATGAGGCTTGGGCTTTGTGAGGAAATCTGCAAGCCGCGAAGTTTCA 324  
Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100  
Db 325 GAAGGGCATGAAATCAACATGGACAAGAAATCGGAGGGGAAAGCCAGCAAGGACTCCGT 384  
Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120  
Db 385 GAGCCACTGGATGGATGACATGAACCGCAGAGCGGTATGCAAGCACATGAAGCGC 444  
Qy 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140  
Db 445 AGCATGGCCGCGCGCTCCAGTTAGCAGACACGCTTTTCTATCATGGAGACTTCGTC 504  
Qy 141 ValValThrAspGlyCysCysSerSerAsnGlyArgLysProArgAlaGlyTyr 160  
Db 505 GTTCGTCTACTGATGGCTGCTGCTCCAGTATGGCGTAGAAGCCGCGGAGCAGGATC 564  
Qy 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyLysArgLeuProGlyArgGln 180  
Db 565 GGCCTTTACTGGGGCCAGGCCATCTTTAAATATAGGCATTTAGACTTCTCGGGCGGAG 624  
Qy 181 ThrAsnGluArgAlaGluLeuHisAlaCysLysAlaLeuGluAlaLysThrGln 200  
Db 625 ACAACCAAGACGGGAATTTCTGAGCTGCAAGCCATTGAACCAAGCAAGACTCAA 684  
Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220

Db 685 AACATCAATAAACTGGTCTCTGTATATACAGACAGATGTTTACGATAAATGGTATACTAAC 744  
Qy 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240  
Db 745 TGGGTTCAAGGTTGGAGAAAATAATGGTGGAAAGACAAGTGCAGGGAAGAGGTTGATCAAC 804  
Qy 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260  
Db 805 AAAGAGACTTTGTGCACTGGAGAGGCTTACCCAGGGATGGACATTCAGTGGATGTCAT 864  
Qy 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280  
Db 865 GTTCCTGTCATTCGGGATTTATAGCAATGAAGAGCTGACAGATTACCCAGAGAAGCA 924  
Qy 281 AlaLysGlnSerGluAsp 286  
Db 925 GCTAAACAATCGGAAGAC 942  
RESULT 3  
ADA05725  
ID ADA05725 standard; cDNA; 965 BP.  
XX AC ADA05725;  
XX DT 06-NOV-2003 (first entry)  
XX DE Human NOV15e encoding cDNA SEQ ID NO:85.  
XX KW human; NOVX; antidiabetic; anorectic; antibacterial; virucide;  
KW immunomodulator; cytostatic; nootropic; neuroprotective;  
KW antiparkinsonian; antilipemic; gene therapy; human disease;  
KW metabolic disorder; diabetes; obesity; infection; cachexia; cancer;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW immune disorder; haematopoietic disorder; dyslipidaemia; gene; ss.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
FT CDS 10..870  
FT FT /\*tag= a  
FT FT /product= "NOV15e"  
XX FN WO2003029424-A2.  
XX 10-APR-2003.  
XX 02-OCT-2002; 2002WO-US031373.  
XX 02-OCT-2001; 2001US-0326483P.  
XX 05-OCT-2001; 2001US-0327435P.  
XX 05-OCT-2001; 2001US-0327449P.  
XX 09-OCT-2001; 2001US-0327917P.  
XX 09-OCT-2001; 2001US-0328029P.  
XX 09-OCT-2001; 2001US-0328044P.  
XX 09-OCT-2001; 2001US-0328056P.  
XX 12-OCT-2001; 2001US-0328849P.  
XX 15-OCT-2001; 2001US-0329414P.  
XX 17-OCT-2001; 2001US-0330142P.  
XX 18-OCT-2001; 2001US-0330309P.  
XX 22-OCT-2001; 2001US-0341058P.  
XX 24-OCT-2001; 2001US-0339266P.  
XX 24-OCT-2001; 2001US-0343629P.  
XX 29-OCT-2001; 2001US-0349575P.  
XX 01-NOV-2001; 2001US-0346357P.  
XX 17-APR-2002; 2002US-0373260P.  
XX 19-APR-2002; 2002US-0373815P.  
XX 19-APR-2002; 2002US-0373817P.  
XX 19-APR-2002; 2002US-0373826P.  
XX 19-APR-2002; 2002US-0373884P.  
XX 22-APR-2002; 2002US-0374977P.  
XX 16-MAY-2002; 2002US-0381037P.  
XX 16-MAY-2002; 2002US-0381038P.



PR 16-MAY-2002; 2002US-0381042P.  
 PR 17-MAY-2002; 2002US-0381642P.  
 PR 28-MAY-2002; 2002US-0383656P.  
 PR 29-MAY-2002; 2002US-0383831P.  
 PR 25-JUN-2002; 2002US-0391335P.  
 PR 01-OCT-2002; 2002US-00262511.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX Smithson G, Millet I, Peyman JA, Kekuda R, Ju J, Li L, Guo X;  
 PI Patturajan M, Spytek KA, Edinger SR, Ellerman K, Malyankar UM;  
 PI Ort T, Gorman L, Zerhusen BD, Anderson DW, Zhong M, Catterton E;  
 PI Ji W, Miller CE, Rastelli L, Stone DJ, Pena CE, Shenoy SG;  
 PI Shimkets RA, Rothenberg MS, Leach MD, Agee ML, Berghs C, Dipippo VA;  
 PI Eisen A, Gangolli EA, Rieger DK, Spaderna SK;  
 XX  
 DR WPI; 2003-381626/36.  
 DR P-PSDB; ADA05726.  
 XX  
 PT New NOVX polypeptides and nucleic acids, useful for diagnosing,  
 PT preventing or treating NOVX-associated disorders, e.g. diabetes, obesity,  
 PT cancer or dyslipidemia, and in chromosome mapping, tissue typing or  
 PT pharmacogenomics.  
 XX  
 PS Claim 20, page 160; 586pp; English.

CC The present invention describes NOVX proteins, where X can be 1 to 55  
 CC (e.g. NOV1). Also described: (1) a composition comprising a polypeptide  
 CC described above and a carrier; (2) a kit comprising, in one or more  
 CC containers, the composition described above; (3) an isolated nucleic acid  
 CC molecule which encodes a NOVX protein of the invention; (4) a vector  
 CC comprising the nucleic acid molecule described above; (5) a cell  
 CC comprising the above vector; (6) an antibody that immunospecifically  
 CC binds to the polypeptide described above; (7) methods for determining the  
 CC presence or amount of the above polypeptide or nucleic acid molecule in a  
 CC sample; (8) methods for determining the presence of or predisposition to a  
 CC disease associated with altered levels of expression of the above  
 CC polypeptide or nucleic acid molecule in a first mammalian subject; (9) a  
 CC method of identifying an agent that binds to the polypeptide described  
 CC above; (10) a method for identifying a potential therapeutic agent for  
 CC use in treating a pathology that is related to an aberrant expression or  
 CC aberrant physiological interactions of the polypeptide; (11) a method of  
 CC screening for a modulator of activity or of latency or predisposition to  
 CC a pathology associated with the polypeptide; (12) a method for modulating  
 CC the activity of the polypeptide described above; (13) methods of treating  
 CC or preventing a pathology associated with the above polypeptide in a  
 CC mammal; and (14) a method for producing the above polypeptide. NOVX  
 CC sequences have antidiabetic, anorectic, antibacterial, virucide,  
 CC immunomodulator, cytostatic, neurotropic, neuroprotective, antiparkinsonian  
 CC and antilipidemic activities, and can be used in gene therapy. The  
 CC polypeptide is useful in manufacturing a medicament for treating a  
 CC syndrome associated with a human disease. The polypeptide or the nucleic  
 CC acid molecule may be used to diagnose, treat or prevent metabolic  
 CC disorders such as diabetes or obesity, infections, cachexia, cancer,  
 CC neurodegenerative disorders such as Alzheimer's disease or Parkinson's  
 CC disease, immune disorders, hematopoietic disorders and various  
 CC dyslipidaemias. The nucleic acids can also be used as hybridisation  
 CC probes, in chromosome mapping, tissue typing, preventive medicine and  
 CC pharmacogenomics. The present sequence encodes a human NOVX protein from  
 CC the present invention.

SQ Sequence 965 BP; 258 A; 210 C; 295 G; 202 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.: 1.28e-125 Length: 965  
 Score: 1522.00 Matches: 281  
 Percent Similarity: 98.95% Conservative: 2  
 Best Local Similarity: 98.25% Mismatches: 3  
 Query Match: 98.45% Indels: 0  
 DE: 7 Gaps: 0

US-10-054-313-1 (1-286) x ADA05725 (1-965)

Qy 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20  
 Db 10 ATGAGCTGGTTTCTGTCTCTGGCCACAGAGTCGCTTGGCCCTTGGCCCTTCCGCCGC 69  
 Qy 21 GlySerArgGlyPheGlyMetPheTrpAlaValArgArgGlyValGlyValPhe 40  
 Db 70 GGCCTCTCGGGTTTCGGGATGTTCTATGCGTGGAGGGGCGCAAGACCGGGGCTTTT 129  
 Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60  
 Db 130 CTGACCTGGGAATGAGTGCAGACAGCAGCGGTGGACCGGTTTCTCTGCTCCAGATTAGAAG 189  
 Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80  
 Db 190 TTTCCACAGAGATGAGCGCTGGCCTTTGTGAGGAATCTCAGACCCCGAAGTTTCA 249  
 Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100  
 Db 250 GAAGGGCATGAAAATCAACATGACATGCAAGATCGAGCGCAAGCCAGACGACCTCCGT 309  
 Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTrpAlaLysHisMetLysPro 120  
 Db 310 GAGCCACTGATGAGATGACATGAAGCGCAGACCGGTATGCAAGACACATGAAGCCG 369  
 Qy 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140  
 Db 370 AGCGTGAGCGCGCGCTCCAGTTAGCAGAGACACGTTTTTCTTACATGGGAGACTTCGTC 429  
 Qy 141 ValValTyrThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGlyLe 160  
 Db 430 GTCCCTTACACTGATGCTGCTGCTCAGTAATGGGGGTAGAAGCGCGCAGCAGGATC 489  
 Qy 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyLeArgLeuProGlyArgGln 180  
 Db 490 GGCCTTTACTGGGGCCAGGCCATCTTTTAATGTAGGCATTAGACTTCCTGGCGCGCAG 549  
 Qy 181 ThrAsnGlnArgAlaGluLeuHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200  
 Db 550 ACAAAACAAAGAGCGGAATTCATGCAAGCTGCAGCCCTGCAAGCCATTGAACAAAGAACTCAA 609  
 Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyLeThrAsn 220  
 Db 610 AACATCATAACTGGTTCTGTATACAGACAGATATGTTTACGATAATGGTATAACTAAC 669  
 Qy 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240  
 Db 670 TGGGTTCAAGGTTGCAAGAAAAATGGCTGGAAGCAAGTGCAGGGAAGAGGTGATCAAC 729  
 Qy 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260  
 Db 730 AAAGAGGACTTTTGGCACTGGAGAGGCTTACCAGGGGGATGGACATTCAGTGGATGCAT 789  
 Qy 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280  
 Db 790 GTTCTCTGTCATTCCGGATTATAGGCAATAGGAAGCTGACAGATTAGCCAGAGAAGGA 849  
 Qy 281 AlalysGlnSerGluAsp 286  
 Db 850 GCTAAACAATCGGAAGAC 867  
 RESULT 4  
 ADC39107  
 ID ADC39107 standard; cDNA; 979 BP.  
 XX  
 AC ADC39107;  
 XX  
 DT 18-DEC-2003 (first entry)  
 XX  
 DE Novel human NOVX polypeptide coding sequence SEQ ID NO: 49.  
 XX  
 KW ds; gene; antidiabetic; cytostatic; immunomodulator; anorectic;  
 KW antilipemic; neurotropic; neuroprotective; immunostimulant;  
 KW antiparkinsonian; anti-Hiv; antiasthmatic; antiinflammatory; hypotensive;



Db 747 TGGGTTCAAGTTGGAGAAATATCGGTGGAGACACAGTGCAGGAAAGAGTGTATCAAC 805  
Qy 241 LysGluAppPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrp----- 258  
Db 807 AAAGAGGACTTCGTGGCACTGGAGAGGCTGACCCAGGGGATGGACATTCAGTGGCGGAGT 866  
Qy 259 -----MetHisValPro 262  
Db 867 ATCTTAATGTTCAIGTCCCA 887

RESULT 5  
ADA05719  
ID ADA05719 standard; cDNA; 631 BP.  
XX  
XX  
XX ADA05719;  
XX  
XX  
XX 06-NOV-2003 (first entry)  
XX Human NOV15b encoding cDNA SEQ ID NO:79.  
XX human; NOVX; antidiabetic; anorectic; antibacterial; virucide;  
KW immunomodulator; cytostatic; nootropic; neuroprotective;  
KW antiparkinsonian; antilipemic; gene therapy; human disease;  
KW metabolic disorder; diabetes; obesity; infection; cachexia; cancer;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW immune disorder; haematopoietic disorder; dyslipidaemia; gene; ss.  
XX Homo sapiens.  
XX  
XX  
XX Key Location/Qualifiers  
XX CDS 2..631  
XX FT /\*tag a  
XX FT /product= "NOV15b"  
XX  
XX WO2003029424-A2.  
XX  
XX 10-APR-2003.  
XX  
XX  
XX 02-OCT-2002; 2002WO-US031373.  
XX  
XX 02-OCT-2001; 2001US-0326483P.  
XX 05-OCT-2001; 2001US-0327435P.  
XX 05-OCT-2001; 2001US-0327449P.  
XX 03-OCT-2001; 2001US-0327917P.  
XX 09-OCT-2001; 2001US-0328029P.  
XX 09-OCT-2001; 2001US-0328044P.  
XX 09-OCT-2001; 2001US-0328056P.  
XX 12-OCT-2001; 2001US-0328849P.  
XX 15-OCT-2001; 2001US-0329414P.  
XX 17-OCT-2001; 2001US-0330142P.  
XX 18-OCT-2001; 2001US-0330309P.  
XX 22-OCT-2001; 2001US-0341058P.  
XX 24-OCT-2001; 2001US-0339266P.  
XX 24-OCT-2001; 2001US-0343629P.  
XX 29-OCT-2001; 2001US-0349575P.  
XX 01-NOV-2001; 2001US-0346357P.  
XX 17-APR-2002; 2002US-0373260P.  
XX 19-APR-2002; 2002US-0373815P.  
XX 19-APR-2002; 2002US-0373817P.  
XX 19-APR-2002; 2002US-0373826P.  
XX 19-APR-2002; 2002US-0373884P.  
XX 16-MAY-2002; 2002US-0381037P.  
XX 16-MAY-2002; 2002US-0381038P.  
XX 16-MAY-2002; 2002US-0381042P.  
XX 17-MAY-2002; 2002US-0381642P.  
XX 28-MAY-2002; 2002US-0383656P.  
XX 29-MAY-2002; 2002US-0383631P.  
XX 25-JUN-2002; 2002US-0391335P.  
XX 01-OCT-2002; 2002US-00262511.  
XX  
XX (CURA-) CURAGEN CORP.

XX Smithson G, Millet I, Peyman JA, Kekuda R, Ju J, Li L, Guo X;  
PI Patturajan M, Spytek KA, Edinger SR, Ellerman K, Malyankar UM;  
PI Ort T, Gorman L, Zernusen BD, Anderson DW, Zhong W, Catterton B;  
PI Ji W, Miller CE, Rastelli L, Stone DJ, Pena CEA, Shenoy SG;  
PI Shimkets RA, Rotherberg ME, Leach MD, Agee ML, Berghs C, Dipippo VA;  
PI Eisen AJ, Gangolli EA, Rieger DK, Spaderna SK;  
XX WPI; 2003-381626/36.  
XX DR P-PSDB; ADA05720.  
XX  
XX New NOVX polypeptides and nucleic acids, useful for diagnosing,  
PT preventing or treating NOVX-associated disorders, e.g. diabetes, obesity,  
PT cancer or dyslipidemia, and in chromosome mapping, tissue typing or  
PT pharmacogenomics.  
XX  
XX Claim 20; Page 159; 586pp; English.  
XX  
XX The present invention describes NOVX proteins, where X can be 1 to 55  
XX (e.g. NOV1). Also described: (1) a composition comprising a polypeptide  
XX described above and a carrier; (2) a kit comprising, in one or more  
XX containers, the composition described above; (3) an isolated nucleic acid  
XX molecule which encodes a NOVX protein of the invention; (4) a vector  
XX comprising the nucleic acid molecule described above; (5) a cell  
XX comprising the above vector; (6) an antibody that immunospecifically  
XX binds to the polypeptide described above; (7) methods for determining the  
XX presence or amount of the above polypeptide or nucleic acid molecule in a  
XX sample; (8) methods for determining the presence of or predisposition to  
XX a disease associated with altered levels of expression of the above  
XX polypeptide or nucleic acid molecule in a first mammalian subject; (9) a  
XX method of identifying an agent that binds to the polypeptide described  
XX above; (10) a method for identifying a potential therapeutic agent for  
XX use in treating a pathology that is related to an aberrant expression or  
XX aberrant physiological interactions of the polypeptide; (11) a method of  
XX screening for a modulator of activity or of latency or predisposition to  
XX a pathology associated with the polypeptide; (12) a method for modulating  
XX the activity of the polypeptide described above; (13) methods of treating  
XX or preventing a pathology associated with the above polypeptide in a  
XX mammal; and (14) a method for producing the above polypeptide. NOVX  
XX sequences have antidiabetic, anorectic, antibacterial, virucide,  
XX immunomodulator, cytostatic, nootropic, neuroprotective, antiparkinsonian  
XX and antilipemic activities, and can be used in gene therapy. The  
XX polypeptide is useful in manufacturing a medicament for treating a  
XX syndrome associated with a human disease. The polypeptide or the nucleic  
XX acid molecule may be used to diagnose, treat or prevent metabolic  
XX disorders such as diabetes or obesity, infections, cachexia, cancer,  
XX neurodegenerative disorders such as Alzheimer's disease or Parkinson's  
XX disease, immune disorders, haematopoietic disorders and various  
XX dyslipidaemias. The nucleic acids can also be used as hybridisation  
XX probes, in chromosome mapping, tissue typing, preventive medicine and  
XX pharmacogenomics. The present sequence encodes a human NOVX protein from  
XX the present invention.  
XX  
XX SQ Sequence 631 BP; 167 A; 140 C; 189 G; 135 T; 0 U; 0 Other;  
  
Alignment Scores:  
Pred. No.: 1.32e-83 Length: 631  
Score: 1045.50 Matches: 201  
Percent Similarity: 70.63% Conservative: 1  
Best Local Similarity: 70.28% Mismatches: 1  
Query Match: 67.63% Indels: 83  
DB: 7 Gaps: 1  
  
US-10-054-313-1 (1-286) x ADA05719 (1-631)  
  
Qy 1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20  
Db 14 ATGAGCTGTTCTTCCTGCTGCGCCACAGAGTCCCTTGGCCGCTTGCCTGCGCGCGC 73  
Qy 21 GlySerArgGlyPheGlyMetPheTrpAlaValArgArgGlyArgLysThrGlyValPhe 40  
Db 74 GGCTCTCGCGGTTTCGGGATGTTCTATGCTGAGAGGGGCGCCAGACCGGGGTCTTT 133

Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60  
Db 134 CTGACCTGGAAATGATGC-----151  
Qy 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80  
Db 151 -----151  
Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGlnSerGluAlaLysProGlyLysArgLewArg 100  
Db 151 -----151  
Qy 101 GluProLeuAspGlyAspGlyHisGlnSerAlaGlnProTyrAlaLysHisMetLysPro 120  
Db 151 -----151  
Qy 121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140  
Db 152 -----AGACACAGTTCCTACATGGGAGACTTCGTC 184  
Qy 141 ValValTyrThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGlyLe 160  
Db 185 GTCGCTACACTGATGGCTGCTCCAGTAATGGCGTAGAAGCGCGCAGCAGGAATC 244  
Qy 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyLeArgLewProGlyArgGln 180  
Db 245 GCGCTTACTGGGGCGCGGCATCCTTTAAATGTAGGCATTAGACTTCCTGGCGGCAG 304  
Qy 181 ThrAsnGlnArgAlaGluLeuHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200  
Db 305 ACAACCAAGAGCGGAATTCATGCGAGCCTGCAAGCCATTGAACAGCAAGACTCAA 364  
Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220  
Db 365 AACATCAATAAACTGGTTCTGTATACACAGCATGTATTAGATAAAATGGTATACTAAC 424  
Qy 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240  
Db 425 TGGGTTCAAGTTGGAAGAAAATGGGTGGAGCAAGTGCAGGGAAGAGGTGTATCAAC 484  
Qy 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260  
Db 485 AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGATGGACATTCAGTGGATGCAT 544  
Qy 261 ValProGlyHisSerGlyPheIleGlyAsnGluAlaLeuAspArgLewAlaArgGluGly 280  
Db 545 GTTCTGTGTCATTGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGGA 604  
Qy 281 AlaLysGlnSerGluAsp 286  
Db 605 GCTAAACAATCGGAAGAC 622

RESULT 6

ADA05717  
ID ADA05717 standard; cDNA; 716 BP.

XX AC ADA05717;

XX XX 06-NOV-2003 (first entry)

DE DE Human NOV15a encoding cDNA SEQ ID NO:77.

XX KW human; NOVX; antidiabetic; anorectic; antibacterial; virucide;

XX KW immunomodulator; cytostatic; nootropic; neuroprotective;

XX KW antiparkinsonian; antilipemic; gene therapy; human disease;

XX KW metabolic disorder; diabetes; obesity; infection; cachexia; cancer;

XX KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;

XX KW immune disorder; haematopoietic disorder; dyslipidaemia; gene; ss.

OS OS Homo sapiens.

XX XX Key Location/Qualifiers

FX FX 10..621

FT FT CDS

FT /\*tag= a  
XX /product= "NOV15a"

PN WO2003029424-A2.

XX 10-APR-2003.

XX XX

PF 02-OCT-2002; 2002WO-US031373.

XX 02-OCT-2001; 2001US-0326483P.

PR 05-OCT-2001; 2001US-0327433P.

PR 05-OCT-2001; 2001US-0327449P.

PR 09-OCT-2001; 2001US-0327917P.

PR 09-OCT-2001; 2001US-0328029P.

PR 09-OCT-2001; 2001US-0328044P.

PR 09-OCT-2001; 2001US-0328056P.

PR 12-OCT-2001; 2001US-0328849P.

PR 15-OCT-2001; 2001US-0329414P.

PR 17-OCT-2001; 2001US-0330142P.

PR 18-OCT-2001; 2001US-0330309P.

PR 22-OCT-2001; 2001US-0341058P.

PR 24-OCT-2001; 2001US-0339266P.

PR 24-OCT-2001; 2001US-0343629P.

PR 29-OCT-2001; 2001US-0349575P.

PR 01-NOV-2001; 2001US-0346357P.

PR 17-APR-2002; 2002US-0373260P.

PR 19-APR-2002; 2002US-0373815P.

PR 19-APR-2002; 2002US-0373817P.

PR 19-APR-2002; 2002US-0373826P.

PR 19-APR-2002; 2002US-0373884P.

PR 22-APR-2002; 2002US-0374977P.

PR 16-MAY-2002; 2002US-0381037P.

PR 16-MAY-2002; 2002US-0381038P.

PR 16-MAY-2002; 2002US-0381042P.

PR 17-MAY-2002; 2002US-0381642P.

PR 28-MAY-2002; 2002US-0383656P.

PR 29-MAY-2002; 2002US-0383831P.

PR 25-JUN-2002; 2002US-0391335P.

PR 01-OCT-2002; 2002US-00262511.

XX (CURA-) CURAGEN CORP.

XX Smithson G, Millet I, Peyman JA, Kekuda R, Ju J, Li L, Guo X;

PI Patturajan M, Spytek KA, Edinger SR, Ellerman K, Malvankar UM;

PI Ort T, Gorman L, Zerhusen BD, Anderson DW, Zhong M, Catterton E;

PI Ji W, Miller CE, Rastelli L, Stone DJ, Pena CE, Shenoy SG;

PI Shimkets RA, Rothenberg ME, Leach MD, Agee ML, Berghs C, Dipippo VA;

PI Eisen AJ, Gangolli EA, Rieger DK, Spaderna SK;

XX WPI; 2003-381626/36.

DR P-PSDB; ADA05718.

XX XX

PT New NOVX polypeptides and nucleic acids, useful for diagnosing,

PT preventing or treating NOVX-associated disorders, e.g. diabetes, obesity,

PT cancer or dyslipidemia, and in chromosome mapping, tissue typing or

PT pharmacogenomics.

XX PS Claim 20; Page 159; 586pp; English.

XX XX The present invention describes NOVX proteins, where X can be 1 to 55

CC (e.g. NOV1). Also described: (1) a composition comprising a polypeptide

CC described above and a carrier; (2) a kit comprising, in one or more

CC containers, the composition described above; (3) an isolated nucleic acid

CC molecule which encodes a NOVX protein of the invention; (4) a vector

CC comprising the nucleic acid molecule described above; (5) a cell

CC comprising the above vector; (6) an antibody that immunospecifically

CC binds to the polypeptide described above; (7) methods for determining the

CC presence or amount of the above polypeptide or nucleic acid molecule in a

CC sample; (8) methods for determining the presence of or predisposition to

CC a disease associated with altered levels of expression of the above

CC polypeptide or nucleic acid molecule in a first mammalian subject; (9) a

CC method of identifying an agent that binds to the polypeptide described

CC above; (10) a method for identifying a potential therapeutic agent for

use in treating a pathology that is related to an aberrant expression or aberrant physiological interactions of the polypeptide; (11) a method of screening for a modulator of activity or of latency or predisposition to a pathology associated with the polypeptide; (12) a method for modulating the activity of the polypeptide described above; (13) methods of treating or preventing a pathology associated with the above polypeptide in a mammal; and (14) a method for producing the above polypeptide. NOVX sequences have anti-diabetic, anorectic, antibacterial, virucide, immunomodulator, cytostatic, neurotropic, neuroprotective, antiparkinsonian and anti-lipemic activities, and can be used in gene therapy. The polypeptide is useful in manufacturing a medicament for treating a syndrome associated with a human disease. The polypeptide or the nucleic acid molecule may be used to diagnose, treat or prevent metabolic disorders such as diabetes or obesity, infections, cachexia, cancer, neurodegenerative disorders such as Alzheimer's disease or Parkinson's disease, immune disorders, haematopoietic disorders and various dyslipidaemias. The nucleic acids can also be used as hybridisation probes, in chromosome mapping, tissue typing, preventive medicine and pharmacogenomics. The present sequence encodes a human NOVX protein from the present invention.

SQ Sequence 716 BP; 185 A; 152 C; 217 G; 162 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.: 1.54e-83 Length: 716  
Score: 1045.50 Matches: 201  
Percent Similarity: 70.63% Conservative: 1  
Best Local Similarity: 70.28% Mismatches: 1  
Query Match: 67.63% Indels: 83  
DB: 7 Gaps: 1

US-10-054-313-1 (1-286) x ADA05717 (1-716)

QY	1	MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg	20
DB	10	ATCAGCTGGTTCTGTTCTCGGCCACAGAGTCGCCCTTGGCGCTTGCCTGCGCGCCG	69
QY	21	GlySerArgGlyPheGlyMetPheTrpAlaValArgArgGlyArgGlyThrGlyValPhe	40
DB	70	GGCTCTCGGGGTTCGGAGTGTCTATGCCGTGAGAGGGCGCGAAGACCGGGGTCCTT	129
QY	41	LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys	60
DB	130	CTGACCTGGATGAGTGC-----	147
QY	61	PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer	80
DB	147	-----	147
QY	81	GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg	100
DB	147	-----	147
QY	101	GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTrpAlaLysHisMetLysPro	120
DB	147	-----	147
QY	121	SerValGluProAlaProProValSerArgAspThrPheSerTrpMetGlyAspPheVal	140
DB	148	-----AGAGACACGTTTCTCTACATGGAGACTTCGTC	180
QY	141	ValValTrpThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGlyIle	160
DB	181	GTCGCTACACTGATGCTGCTCCAGTAATGGCGTAGAAGCGCGGAGCAGCAATC	240
QY	161	GlyValTrpTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuPheGlyArgGln	180
DB	241	GGCGTTTACGGGGCGCGGCCCATCTCTTTAAAGTAGAGCAATTAGACTTCTGGCGGCGAG	300
QY	181	ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln	200
DB	301	ACAAACCAAGAGCGGAAATTCATGCGCTGCAAGCCATTGAACAAGCAAGACTCAA	360

QY	201	AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn	220
DB	361	AACATCATTAACCTGGTCTGTATACAGACAGTATGTTTACGATAAATGGTAACTAAC	420
QY	221	TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn	240
DB	421	TGGGTTCAAGGTTTGGAGAAAAAATGGGTGGAGACAAGTCCAGGAAAGAGGTGATCAAC	480
QY	241	LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis	260
DB	481	AAAGAGACATTGTGGCCTGGAGAGCTTACCCAGGGGATGGACATTCAGTGGATGCAT	540
QY	261	ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly	280
DB	541	GTTCTCTGCTCATTCGGGATTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGGA	600
QY	281	AlaLysGlnSerGluAsp	286
DB	601	GCTAAACAATCGGAAGAC	618
RESULT 7			
ADA05721	ID	ADA05721 standard; cDNA; 586 BP.	
XX	AC	ADA05721;	
XX	DT	06-NOV-2003 (first entry)	
XX	DE	Human NOV15c encoding cDNA SEQ ID NO:81.	
XX	KW	human; NOVX; antidiabetic; anorectic; antibacterial; virucide; immunomodulator; cytostatic; neurotropic; neuroprotective; antiparkinsonian; antilipemic; gene therapy; human disease; metabolic disorder; diabetes; obesity; infection; cancer; neurodegenerative disorder; Alzheimer's disease; Parkinson's disease; immune disorder; haematopoietic disorder; dyslipidaemia; gene; ss.	
XX	OS	Homo sapiens.	
XX	FH	Key	Location/Qualifiers
FT	CDS	2..586	/*tag= a
FT		/product= "NOV15c"	
XX	PN	WO2003029424-A2.	
XX	PD	10-APR-2003.	
XX	PF	02-OCT-2002; 2002WO-US031373.	
XX	PR	02-OCT-2001; 2001US-0326483P.	
PR	PR	05-OCT-2001; 2001US-0327435P.	
PR	PR	05-OCT-2001; 2001US-0327449P.	
PR	PR	09-OCT-2001; 2001US-0327917P.	
PR	PR	09-OCT-2001; 2001US-0328029P.	
PR	PR	09-OCT-2001; 2001US-0328044P.	
PR	PR	12-OCT-2001; 2001US-0328849P.	
PR	PR	15-OCT-2001; 2001US-0329414P.	
PR	PR	17-OCT-2001; 2001US-0330142P.	
PR	PR	18-OCT-2001; 2001US-0330309P.	
PR	PR	22-OCT-2001; 2001US-0341038P.	
PR	PR	24-OCT-2001; 2001US-0339266P.	
PR	PR	29-OCT-2001; 2001US-0343629P.	
PR	PR	29-OCT-2001; 2001US-0349575P.	
PR	PR	01-NOV-2001; 2001US-0346357P.	
PR	PR	17-APR-2002; 2002US-0373260P.	
PR	PR	19-APR-2002; 2002US-0373815P.	
PR	PR	19-APR-2002; 2002US-0373817P.	
PR	PR	19-APR-2002; 2002US-0373826P.	
PR	PR	19-APR-2002; 2002US-0373884P.	
PR	PR	22-APR-2002; 2002US-0374977P.	
PR	PR	16-MAY-2002; 2002US-0381037P.	



XX 23-AUG-2001.  
XX 20-FEB-2001; 2001WO-US005171.  
XX 17-FEB-2000; 2000US-0183319P.  
XX 16-MAR-2000; 2000US-0189862P.  
XX 25-MAY-2000; 2000US-0207454P.  
XX 09-JUN-2000; 2000US-0211314P.  
XX 18-JUL-2000; 2000US-0219007P.  
XX 13-DEC-2000; 2000US-0255281P.  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX Schlegel R, Endege WO, Monahan JE;  
XX WPI; 2001-662795/76.  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
XX prostate cells and correlating with presence of prostate cancer, useful  
XX for detecting presence of prostate cancer, stage of prostate cancer.  
XX Claim 1; Page 9051; 11750pp; English.  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
XX specification or its complement. (I) is useful for: (a) assessing whether  
XX a patient is afflicted with prostate cancer; (b) monitoring the  
XX progression of prostate cancer in a patient; (c) assessing the efficacy  
XX of a test compound to inhibit prostate cancer in a patient; (d) assessing  
XX the efficacy of a therapy for inhibiting prostate cancer in a patient;  
XX (e) selecting a composition for inhibiting prostate cancer in a patient;  
XX (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
XX determining whether prostate cancer has metastasized in a patient; (h)  
XX assessing the aggressiveness or indolence of prostate cancer in a patient  
XX ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker  
XX  
XX SQ Sequence 545 BP; 136 A; 129 C; 178 G; 102 T; 0 U; 0 Other;  
Alignment Scores:  
Pred. No.: 3.15e-71 Length: 545  
Score: 905.00 Matches: 167  
Percent Similarity: 94.44% Conservative: 3  
Best Local Similarity: 92.78% Mismatches: 10  
Query Match: 58.54% Indels: 0  
DB: Gaps: 0  
US-10-054-313-1 (1-286) x ABV45787 (1-545)  
QY 13 LeuAlaValProCysArgArgGlySerArgGlyPheGlyMetPheTyrAlaValArg 32  
Db 4 ATTGGAGCTCCACCGCGGTGGCGCGAGGTGCGGGTTCGGGATGTTCTATGCCGTGAGG 63  
QY 33 ArgGlyArgLysThrGlyValPheLeuThrTrpAsnGluCysArgAlaGlnValAspArg 52  
Db 64 AGGGCCCGCAGACTGGGGTCTTCTGACCTGGAATGAGTGCAGACACAGGTGGACCGG 123  
QY 53 PheProAlaAlaArgPheLysLysPheAlaThrGluAspGluAlaTrpAlaPheValArg 72  
Db 124 TTTCCTCTCCGAGATTAGAAATTGTCACAGAGATGAGCCCTGGGCTTTGTCAGG 183  
QY 73 LysSerAlaSerProGluValSerGluGlyHisGluAsnGlnHisGlyGlnGluSerGlu 92  
Db 184 AATCTGCAAGCCCGGAGTTTCAGAGGCGCATGAAATCAATGACACAGAGATCGGAG 243  
QY 93 AlaLysProGlyLysArgLeuArgGluProLeuAspGlyAspGlyHisGluSerAlaGln 112  
Db 244 GCGAAGCCGACGACGACTCCGTGAGCCACTGGATGGAGATGAGCATGAAAGCGCAGAG 303  
QY 113 ProTyrAlaLysHisMetLysProSerValGluProAlaProValSerArgAspThr 132  
Db 304 CCGTATGCAAGACATGAAAGCCGAGCGTGGAGCCGCGCTTCAGTTAGCAGAGACG 363

QY 133 PheSerTyrMetGlyAspPheValValValTyrThrAspGlyCysCysSerSerAsnGly 152  
Db 364 TTTTCTCTACATGGGAGACTTCGTCGTCGTCTACACTGATGGCTGCTCCAGTAATGGG 423  
QY 153 ArgArgLysProArgAlaGlyIleGlyValTyrTrpGlyProGlyHisProLeuAsnVal 172  
Db 424 CGTAGAAGCGCGGAGCAGGAGTTCGGGTTTACTGGGGCCCGGCCATCTCTTTAAATGTA 483  
QY 173 GlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGluIleHisAlaAlaCysLys 192  
Db 484 GGCATTAGACTTCTGGCGGCGAGACAAACCAAGAGCGGAAATTCATGCAACTGCAAA 543  
RESULT 9  
ADA05723  
ID ADA05723 standard; cDNA; 457 BP.  
XX  
AC ADA05723;  
XX  
DT 06-NOV-2003 (first entry)  
XX  
DE Human NOV15d encoding cDNA SEQ ID NO:83.  
XX  
KW human; NOVX; antidiabetic; anorectic; antibacterial; virucide;  
KW immunomodulator; cytostatic; nootropic; neuroprotective;  
KW antiparkinsonian; antilipemic; gene therapy; human disease;  
KW metabolic disorder; diabetes; obesity; infection; cachexia; cancer;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW immune disorder; haematopoietic disorder; dyslipidaemia; gene; ss.  
XX  
OS Homo sapiens.  
XX  
Key Location/Qualifiers  
CDS 2..457  
FT /\*tag= a  
FT /product= "NOV15d"  
XX  
PW WO2003029424-A2.  
XX  
PD 10-APR-2003.  
XX  
XX 02-OCT-2002; 2002WO-US031373.  
XX  
XX 02-OCT-2001; 2001US-0326483P.  
XX 05-OCT-2001; 2001US-0327435P.  
XX 05-OCT-2001; 2001US-0327449P.  
XX 09-OCT-2001; 2001US-0327917P.  
XX 09-OCT-2001; 2001US-0328029P.  
XX 09-OCT-2001; 2001US-0328044P.  
XX 09-OCT-2001; 2001US-0328056P.  
XX 12-OCT-2001; 2001US-0328849P.  
XX 15-OCT-2001; 2001US-0329414P.  
XX 17-OCT-2001; 2001US-0330142P.  
XX 18-OCT-2001; 2001US-0330309P.  
XX 22-OCT-2001; 2001US-0341058P.  
XX 24-OCT-2001; 2001US-0343629P.  
XX 29-OCT-2001; 2001US-0349575P.  
XX 01-NOV-2001; 2001US-0346357P.  
XX 17-APR-2002; 2002US-0373260P.  
XX 19-APR-2002; 2002US-0373815P.  
XX 19-APR-2002; 2002US-0373817P.  
XX 19-APR-2002; 2002US-0373826P.  
XX 22-APR-2002; 2002US-0374977P.  
XX 16-MAY-2002; 2002US-0381037P.  
XX 16-MAY-2002; 2002US-0381038P.  
XX 17-MAY-2002; 2002US-0381042P.  
XX 28-MAY-2002; 2002US-0381642P.  
XX 29-MAY-2002; 2002US-0383656P.  
XX 25-JUN-2002; 2002US-0383831P.  
XX 01-OCT-2002; 2002US-0391335P.  
XX 01-OCT-2002; 2002US-00362511.  
XX

(CURA-) CURAGEN CORP.

PA Smithson G, Millet I, Peyman JA, Kekuda R, Ju J, Li L, Guo X;  
 PI Patcurayan M, Spytek KA, Edinger SR, Ellerman K, Malyankar UM;  
 PI Ort T, Gorman L, Zerhusen BD, Anderson DW, Zhong M, Catterton E;  
 PI Ji W, Miller CE, Rastelli L, Stone DJ, Pena CEA, Shenoy SG;  
 PI Shinkets RA, Rotherberg ME, Leach WD, Agee ML, Berghs C, Dipippo VA;  
 PI Eisen AJ, Gangolli EA, Rieger DK, Spaderna SK;  
 XX  
 DR WPI: 2003-381626/36.  
 DR P-PSDB; ADA05724.  
 XX  
 PT New NOVX polypeptides and nucleic acids, useful for diagnosing,  
 PT preventing or treating NOVX-associated disorders, e.g. diabetes, obesity,  
 PT cancer or dyslipidemia, and in chromosome mapping, tissue typing or  
 PT pharmacogenomics.  
 PT  
 XX  
 PS Claim 20; Page 160; 586pp; English.  
 XX  
 CC The present invention describes NOVX proteins, where X can be 1 to 55  
 CC (e.g. NOV1). Also described: (1) a composition comprising a polypeptide  
 CC described above and a carrier; (2) a kit comprising, in one or more  
 CC containers, the composition described above; (3) an isolated nucleic acid  
 CC molecule which encodes a NOVX protein of the invention; (4) a vector  
 CC comprising the nucleic acid molecule described above; (5) a cell  
 CC comprising the above vector; (6) an antibody that immunospecifically  
 CC binds to the polypeptide described above; (7) methods for determining the  
 CC presence or amount of the above polypeptide or nucleic acid molecule in a  
 CC sample; (8) methods for determining the presence of or predisposition to  
 CC a disease associated with altered levels of expression of the above  
 CC polypeptide or nucleic acid molecule in a first mammalian subject; (9) a  
 CC method of identifying an agent that binds to the polypeptide described  
 CC above; (10) a method for identifying a potential therapeutic agent for  
 CC use in treating a pathology that is related to an aberrant expression or  
 CC aberrant physiological interactions of the polypeptide; (11) a method of  
 CC screening for a modulator of activity or of latency or predisposition to  
 CC a pathology associated with the polypeptide; (12) a method for modulating  
 CC the activity of the polypeptide described above; (13) methods of treating  
 CC or preventing a pathology associated with the above polypeptide in a  
 CC mammal; and (14) a method for producing the above polypeptide. NOVX  
 CC sequences have antidiabetic, anorectic, antibacterial, virucide,  
 CC immunomodulator, cytostatic, nootropic, neuroprotective, antiparkinsonian  
 CC and antilipidemic activities, and can be used in gene therapy. The  
 CC polypeptide is useful in manufacturing a medicament for treating a  
 CC syndrome associated with a human disease. The polypeptide or the nucleic  
 CC acid molecule may be used to diagnose, treat or prevent metabolic  
 CC disorders such as diabetes or obesity, infections, cachexia, cancer,  
 CC neurodegenerative disorders such as Alzheimer's disease or Parkinson's  
 CC disease, immune disorders, haematopoietic disorders and various  
 CC dyslipidaemias. The nucleic acids can also be used as hybridisation  
 CC probes, in chromosome mapping, tissue typing, preventive medicine and  
 CC pharmacogenomics. The present sequence encodes a human NOVX protein from  
 CC the present invention.  
 XX  
 SQ Sequence 457 BP; 139 A; 89 C; 134 G; 95 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 2,64e-61 Length: 457  
 Score: 792.00 Matches: 145  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 99.32% Mismatches: 0  
 Query Match: 51.23% Indels: 0  
 DB: 7 Gaps: 0

US-10-054-313-1 (1-286) x ADA05723 (1-457)

QY 137 GlyAspPheValValValValThrAspGlyCysCysSerSerAsnGlyArgGlyPro 156  
 DB 11 GGAGATTCTGCTGCTGTACATGATGCTGCTCCAGTAATGGCGTGAAGCGG 70  
 QY 157 ArgAlaGlyIleGlyValValThrGlyProGlyHisProLeuAsnValGlyIleArgLeu 176  
 DB 71 CGAGCAGGAATCGCGCTTACTTGGGGCCGGCCATCCCTTAAATGTAGCATTTAGACTT 130

QY 177 ProGlyArgGlnThrAsnGlnArgAlaGluIleHisAlaAlaCysLysValIleGluGln 196  
 DB 131 CCTGGGGCGGAGACAACCAAGAGCGGAAATTCATGACCTTGAAGCCATTTGACAA 190  
 QY 197 AlalyThrGlnAsnIleAsnLysLeuValLeuValThrAspSerMetPheThrIleAsn 216  
 DB 191 GCAAGAGACTCAAAACATCAATAAAGCTGTTCTGTATATACACAGACAGTAGTATGTTACGATAAAT 250  
 QY 217 GlyIleThrAsnTrpValGlnGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLys 236  
 DB 251 GGTATTAACCTAAGTGGGTTCAAGGTTGAAGAAAAATGGGTGGAGACAAGTGCAGGAAA 310  
 QY 237 GluValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIle 256  
 DB 311 GAGTGATCAACAAGAGAGACTTTTGTGGCACTGGAGAGGCTTATCCAGGGGATGGACATT 370  
 QY 257 GlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGlnGluAlaAspArgLeu 276  
 DB 371 CAGTGGATGATGTTCTGTCATTCGGGATTTATAGGCAATGAAGAAGCTGACAGATTA 430  
 QY 277 AlaArgGluGlyAlaLys 282  
 DB 431 GCCAGAGAGAGAGCTTAA 448

RESULT 10  
 ABV15988  
 ID ABV15988 standard; cDNA; 691 BP.  
 AC ABV15988;  
 XX 13-SEP-2002 (first entry)  
 DT Human prostate expression marker cDNA 15979.  
 DE Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW pharmacogenomic marker; gene; ss.  
 XX Homo sapiens.  
 OS WO200160860-A2.  
 FN 23-AUG-2001.  
 PD 20-FEB-2001; 2001WO-US0005171.  
 PF 17-FEB-2000; 2000US-0183319P.  
 PR 16-MAR-2000; 2000US-0189862P.  
 PR 25-MAY-2000; 2000US-0207454P.  
 PR 09-JUN-2000; 2000US-0211314P.  
 PR 18-JUL-2000; 2000US-0219007P.  
 PR 13-DEC-2000; 2000US-0255281P.  
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 FA Schlegel R, Endege WO, Monahan JE;  
 FI WPI; 2001-662795/76.  
 XX Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer, useful  
 PT for detecting presence of prostate cancer, stage of prostate cancer.  
 XX  
 PS Claim 1; Page 2678; 11750pp; English.  
 XX The invention relates to an isolated nucleic acid molecule (I) comprising  
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
 CC specification or its complement. (I) is useful for: (a) assessing whether  
 CC a patient is afflicted with prostate cancer; (b) monitoring the  
 CC progression of prostate cancer in a patient; (c) assessing the efficacy  
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing  
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;



CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
CC determining whether prostate cancer has metastasized in a patient; (h)  
CC assessing the aggressiveness or incidence of prostate cancer in a patient  
CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker  
XX  
SQ Sequence 691 BP; 186 A; 155 C; 199 G; 144 T; 0 U; 7 Other;

Alignment Scores:  
Pred. No.: 5,86e-50 Length: 691  
Score: 666.50 Matches: 159  
Percent Similarity: 73.82% Conservative: 13  
Best Local Similarity: 68.24% Mismatches: 34  
Query Match: 43.11% Indels: 28  
DB: 5 Gaps: 8

US-10-054-313-1 (1-286) x ABV15988 (1-691)

QY 24 GlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPheLeuThrTrp 43  
Db 3 GGGTTCGGGATGTTCTATGCGGTGAGAGGGGCCGCAAGACTGGGGTCTTTCTGACCTGG 62  
QY 44 AsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLysPheAlaThr 63  
Db 63 AATGAGTGCAGACACAGGTGGACCGGTTCTGCTGCCAGATTGAGAGTTTGCACA 122  
QY 64 GluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHis 83  
Db 123 GAGGATGAGCGCTGGGCGCTTGTTCAGAAATCTGCAAGCCCGGAGATTTCAGAGGGCAT 182  
QY 84 GluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGluProLeu 103  
Db 183 GAAATCAATCATGACACAGANTCGGAGCGGAAAGCCAGCAGGATCTCGTAGCCACGTG 242  
QY 104 AspGlyAspGlyHisGluSerAlaGln-ProTyrAlaLysHis-MetLysProSerVal- 122  
Db 243 GATGGAGATGCATGAAGCGCAGACCGCGTATGCAAGCACCATTGAAGCCGAGCCGT 302  
QY 123 --GluProAlaProValSerArgAspThr-PheSerTyrMet-GlyAspPheVal-V 141  
Db 303 GAGCGCGCGCTTCAGTTAGCAGACACCGCTTTTCTTACATGGGGAGACTTCGTTGG 362  
QY 141 alValTyrThrAspGly-CysCysSerSerAsnGlyArgLys-----ProArgAla 158  
Db 363 TCGTCTACATGATGGCTTCTGCTCCAGTAATGGCGGTANGAGGCCCGCGGAGCAG 422  
QY 159 GlylleGlyValTyrTrp----GlyProGlyHisProLeuAsn---ValGlylleArgLeu 176  
Db 423 GGAATCGCGCTTTACTGGGGGGCCCGCGCCATCTTTAAATGGTAGGCCATTAAAGACTT 482  
QY 177 Pro-----GlyArg-GlnThrAsnGlnArgAlaGluIleHisAlaAla-----CysLys 192  
Db 483 CCTTGGCGCGCAGGACCAACCCAAAGAGGCGGGNMAATTTATGCCANGCCCTTGCAA 542  
QY 192 sAlaIleGluGlnAlaLysThrGlnAsnIleAsnLys----- 204  
Db 543 AAGCCCATTTTGAACCAAGCCAAAGGAGCTTCAAAAACCATTCAAATAAAACCTGGGG 602  
QY 205 -LeuValLeuTyrThrAspSerMetPheThrIle-----AsnGlylleThrAsnTr 221  
Db 603 TTCTCTGTTATTCAGACACAA-GTNTTGTGTTTACCGAATAAATGGGNTTTAAACTT 561  
QY 221 pValGlnGlyTyrPheLysAsnGly 229  
Db 662 AACCTTGGGTTTCAANGGTTNGGG 686

RESULT 11

ADD34573

ID ADD34573 standard; DNA; 519 BP.

XX AC

XX ADD34573;

DT 15-JAN-2004 (first entry)

XX

Mouse mitochondrial DNA sequence SEQ ID NO:2351.

ds; mouse; array; mitochondrial; hybridisation; energy-metabolism;  
mitochondrial disease; oxidative phosphorylation dysfunction;  
oxidative stress; apoptosis; aging.

Mus musculus.

WO2003020220-A2.

13-MAR-2003.

30-AUG-2002; 2002WO-US027866.

30-AUG-2001; 2001US-0316323P.

31-AUG-2001; 2001CA-02356540.

(UYEM-) UNIV EMORY.

Wallace DC, Levy S, Kerstann K, Procaccio V;

WPI; 2003-300821/29.

Array containing probes for genes involved in mitochondrial biology,  
useful for determining mitochondrial biology gene expression profiles for  
use in diagnosing pathologies and identifying biochemical pathways.

Claim 2; SEQ ID NO 2351; 201pp; English.

The invention relates to a novel array comprising at least two isolated  
nucleotide molecules, each molecule having a sequence capable of uniquely  
hybridising to a nucleic acid molecule which is an expression product of  
a gene involved in mitochondrial biology. The array comprises two or more  
isolated nucleic acid molecules or spots, each molecule having a sequence  
chosen from sequence of 994 human probes and 2046 mouse probes. An array  
of the invention is useful for determining an expression profile of a  
mouse or human sample containing nucleic acid, by contacting the array  
with the sample under conditions allowing selective hybridisation, and  
measuring hybridisation of nucleic acid in the sample to the array to  
produce an expression profile. The array is also useful for determining  
an expression profile of a first labelled sample containing nucleic acid  
relative to a second, differently labelled sample containing nucleic acid  
CC The second sample is a reference or a standard. An array is useful  
CC for determining an expression profile diagnostic of an energy-metabolism-  
CC related physiological condition. An array of the invention is useful for  
CC determining mitochondrial biology gene expression profiles of organisms,  
CC such as human, mice and closely related species, tissue and organs of  
CC diagnostic of energy metabolism-related physiological conditions,  
CC diagnosing such physiological conditions, identifying biochemical  
CC pathways, genes, and mutations involved in such physiological conditions,  
CC identifying therapeutic agents useful for preventing and/or treating such  
CC physiological conditions, evaluating and/or monitoring the efficacy of  
CC such therapies, and creating and identifying animal models of human  
CC energy metabolism-related physiological conditions. An array is also  
CC useful for defining expression signatures or profiles for mitochondrial  
CC diseases, as well as distinguishing clinical disorders that result from  
CC oxidative phosphorylation (OXPHOS) dysfunction, oxidative stress,  
CC apoptosis and aging. An array of the invention contains probes of genes  
CC not previously recognised to participate in mitochondrial biology. The  
CC sequences shown in ADD33224-ADD33260 represent murine mitochondrial DNA  
CC clones used to make the probes of the invention. Some sequences are not  
CC present, these are SEQ ID NO's 295, 1174, 1213, 1700, 1728, 1730, 1905,  
CC 1906, 2408 and 2643.

Seq Sequence 519 BP; 137 A; 120 C; 163 G; 99 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

Length:

Matches:

Conservative:

Mismatches:

Indels:

519

112

11

7

21



Db 582 GGTCACTAGGATTGTGGCCAAATGAAGAGCGGCACAGATGTCACCGGAGGAGCGAAG 523  
QY 283 GlnSerGluasp 286  
Db 522 CAGTCTGAGGAC 511

## RESULT 13

AAH30241  
ID AAH30241 standard; cDNA; 389 BP.

XX AAH30241;

XX 27-JUL-2001 (first entry)

XX Human colon cancer cell line Kml2L4-A cDNA library derived sequence #175.

XX Human; diagnosis; colon cancer; cancer; malignant; chromosome mapping;  
KW detection; colon cancer cell line Kml2L4-A; 88.  
XX Homo sapiens.

OS WO200018916-A2.

XX 06-APR-2000.

XX 23-SEP-1999; 99WO-US022226.

XX 28-SEP-1998; 98US-0102161P.

XX 28-SEP-1998; 98US-0102180P.

XX 29-SEP-1998; 98US-0102380P.

XX 08-OCT-1998; 98US-0103815P.

XX 27-OCT-1998; 98US-0105877P.

XX (CHIR ) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;

PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;

PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;

PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;

XX WPI; 2000-293155/25.

XX Polynucleotide library comprising 1079 defined sequences, useful in the  
PT form of an array to detect cancer or susceptibility to cancer.  
XX Claim 1; Page 233; 502pp; English.

XX The present invention describes a library of polynucleotides comprising  
CC 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described  
CC are: (1) an isolated polynucleotide (I) having at least 90% identity to  
CC one of the 1079 sequences; (2) a recombinant host cell containing (I);  
CC (3) an isolated polypeptide (II) encoded by (I); (4) an antibody that  
CC specifically binds to (II); (5) a vector comprising (I); and (6) a method  
CC of detecting differentially expressed genes correlated with a cancerous  
CC state of a mammalian cell comprising detecting a gene product encoded by  
CC 65 of the 1079 sequences given in the specification. The polynucleotides  
CC are used to monitor patients having (or susceptible) to cancer to detect  
CC potentially malignant events at a molecular level before they are  
CC detectable at a gross morphological level. The polynucleotides are also  
CC useful for monitoring the efficacy of various therapies and preventive  
CC interventions. Polynucleotide probes based on the disclosed sequences are  
CC useful for chromosome mapping and detection of transcription levels. The  
CC 1079 polynucleotide sequences were derived from a human colon cancer cell  
CC line Kml2L4-A cDNA library

SQ Sequence 389 BP; 78 A; 101 C; 130 G; 80 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 2,93e-36 Length: 389  
Score: 508.50 Matches: 104  
Percent Similarity: 56.15% Conservative: 1

Best Local Similarity: 55.61% Mismatches: 0  
Query Match: 32.89% Indels: 82  
DB: 3 Gaps: 1

US-10-054-313-1 (1-286) x AAH30241 (1-389)

QY 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20  
Db 74 ATGAGCTGGCTTCTGCTTCTGGCCACAGAGTCGCTTGGCCGCTTGGCCGCGC 133  
QY 21 GlysSerArgGlyPheGlyMetPheTrpAlaValArgGlyValArgGlyThrGlyValPhe 40  
Db 134 GGCTCTCGCGGTTGGGATGTTCTATGCCGTGAGGAGGGCCGACAGCCGGGCTTT 193  
QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArg-PheLysly 60  
Db 194 CTGACCTGGAATGAGTGCAGAGCACAGGTGCGCGGTTCTCTGCTCCAGACTTC 248

QY 60 sPheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValse 80  
Db 248 ----- 248

QY 80 rGluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuAr 100  
Db 248 ----- 248

QY 100 gGluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPr 120  
Db 248 ----- 248

QY 120 oSerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVa 140  
Db 249 -----GT 250

QY 140 lValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIl 160  
Db 251 CGTCGCTACACTGATGGCTGCTGCCAGTAATGGCGGTAGAGCCGCGACGAGGAAT 310

QY 160 eGlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGl 180  
Db 311 CGCGTTTACTGGGGCCGGCCATCTTTAAATGATGAGCATTAGACTTCCTGGCGGCA 370

QY 180 nThrAsnGlnArgAlaGlu 186

Db 371 GACAAACCAAGAGCGGAA 389

## RESULT 14

ACH42294

ID ACH42294 standard; cDNA; 436 BP.

XX ACH42294;

XX 13-OCT-2003 (first entry)

XX Human foetal brain cDNA #3661.

XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;  
KW genome mapping; biodiversity; genetic disorder.  
XX Homo sapiens.

OS US2003073623-A1.

PN 17-APR-2003.

XX 30-JUL-2001; 2001US-00918995.

XX 30-JUL-2001; 2001US-00918995.

XX (DRNA/) DRMANAC R T.

PA (LABA/) LABAT I.

PA (STAC/) STACHE-CRAIN B.

PA (DICK/) DICKSON M C.





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 22, 2004, 22:32:11 ; Search time 70 Seconds  
(without alignments)  
2267.371 Million cell updates/sec

Title: US-10-054-313-1

Perfect score: 1546

Sequence: 1 MSMLFLAHRVALAALPCRR.....FIGNEADRLREGAKQSD 286

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US10054313/runat 22032004 140416 28103/app query.fasta\_1.455  
-DB=Issued\_Patents\_NA -QWMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFWT=ppc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10054313 @CGN\_1 1.44 @runat 22032004 140416 28103 -NCFU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	272	17.6	1326	US-09-540-236-1702	Sequence 1702, Ap
2	272	17.6	119211	US-09-596-002-40	Sequence 40, Appl
3	268	17.3	231	US-09-621-976-14380	Sequence 14380, A
4	241	15.6	1275	US-09-252-991A-13638	Sequence 13638, A
5	236.5	15.3	1830121	US-09-557-884-1	Sequence 1, Appl
6	236.5	15.3	1830121	US-09-643-990A-1	Sequence 1, Appl
7	235.5	15.2	801	US-09-199-637A-154	Sequence 154, App
8	235.5	15.2	2048	US-09-199-637A-136	Sequence 136, App
9	233.5	15.1	570	US-09-252-991A-13802	Sequence 13802, A
10	217	14.0	1509	US-09-328-352-1765	Sequence 1765, Ap
11	208.5	13.5	603	US-09-489-039A-5519	Sequence 5519, Ap
12	194.5	12.6	489	US-09-543-681A-2007	Sequence 2007, Ap

13	174	11.3	501	1	US-08-163-181-3	Sequence 3, Appl
14	174	11.3	501	1	US-08-465-181-3	Sequence 3, Appl
15	174	11.3	552	1	US-08-163-181-4	Sequence 4, Appl
16	174	11.3	552	1	US-08-465-181-4	Sequence 4, Appl
17	160	10.3	8088	4	US-09-315-127-4	Sequence 4, Appl
18	160	10.3	8535	3	US-08-716-351A-1	Sequence 1, Appl
19	157	10.2	8060	3	US-08-766-528-1	Sequence 1, Appl
20	156	10.1	8202	1	US-08-258-420-13	Sequence 13, Appl
21	154	10.0	2037	4	US-09-397-955C-20	Sequence 20, Appl
22	154	10.0	2037	4	US-09-397-955C-21	Sequence 21, Appl
23	154	10.0	2336	4	US-09-397-955C-18	Sequence 18, Appl
24	154	10.0	3612	4	US-09-265-013-3	Sequence 3, Appl
25	154	10.0	7308	3	US-09-011-745-3	Sequence 3, Appl
26	154	10.0	7308	3	US-09-011-745-4	Sequence 4, Appl
27	154	10.0	7616	3	US-09-011-745-2	Sequence 2, Appl
28	154	10.0	8322	3	US-08-850-961-1	Sequence 1, Appl
29	154	10.0	8322	4	US-09-479-778-1	Sequence 1, Appl
30	154	10.0	8322	4	US-09-309-572-11	Sequence 11, Appl
31	154	10.0	8322	4	US-09-315-127-1	Sequence 1, Appl
32	154	10.0	8322	4	US-09-265-013-1	Sequence 1, Appl
33	154	10.0	8322	4	US-09-554-572-25	Sequence 25, Appl
34	154	10.0	8322	4	US-09-718-096-11	Sequence 11, Appl
35	153	9.9	8363	2	US-08-929-967-6	Sequence 6, Appl
36	152	9.8	8323	1	US-08-110-300A-8	Sequence 8, Appl
37	152	9.8	8323	2	US-08-886-642-8	Sequence 8, Appl
38	152	9.8	8323	4	US-09-433-322B-2	Sequence 2, Appl
39	152	9.8	8323	5	PCT-US93-08041-8	Sequence 8, Appl
40	152	9.8	10367	1	US-08-110-300A-9	Sequence 9, Appl
41	152	9.8	10367	2	US-08-886-642-9	Sequence 9, Appl
42	152	9.8	10367	5	PCT-US93-08041-9	Sequence 9, Appl
43	149	9.6	1953	4	US-09-397-955C-19	Sequence 19, Appl
44	149	9.6	8132	3	US-08-766-528-3	Sequence 3, Appl
45	144	9.3	8655	3	US-09-075-272-1	Sequence 1, Appl

#### ALIGNMENTS

RESULT 1  
US-09-540-236-1702  
; Sequence 1702, Application US/09540236  
; Patent No. 6673910  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CAT  
; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2005-001  
; CURRENT APPLICATION NUMBER: US/09/540,236  
; CURRENT FILING DATE: 2000-04-04  
; NUMBER OF SEQ ID NOS: 3840  
; SEQ ID NO 1702  
; LENGTH: 1326  
; TYPE: DNA  
; ORGANISM: M.catarrhalis  
US-09-540-236-1702

Alignment Scores:  
Pred. No.: 1,26e-20 Length: 1326  
Score: 22.00 Matches: 62  
Percent Similarity: 52.23% Conservative: 20  
Best Local Similarity: 39.49% Mismatches: 51  
Query Match: 17.59% Indels: 24  
DB: 4 Gaps: 5

US-10-054-313-1 (1-286) x US-09-540-236-1702 (1-1326)

Qy 136 MetGlyAspPheValValValTyrThrAspGlyCysSerSerHandgLyArgArgLyLys 155  
Db 28 ATGAGGCAACTTAATATCGCTTATACGCGGCTGTAAGGCAATGGAACAAGC 87  
Qy 156 ProArgAla---GlyIleGlyValTyr----- 163  
Db 88 GTATCTGCAGCGGCTGGGGCGGTGATTTTCATTATTTCATGGCGATGCGGCATCTG 147

QY 164 TrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGlnThrAsnGln 183  
 Db 148 TGGGGC---GGTGAGCCT-----GATACGACCAATAAT 177  
 QY 184 ArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsn 203  
 Db 178 CGCATGGAATTGATGGCTGCCATCAGAGCTTTGGAGGCAACGCTGGCAG-----ATT 231  
 QY 204 LysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGln 223  
 Db 232 CCTCTGCACTTTGACAGATTGAGCTATGTTAAAGATGGCATAACTCAGTGGATTGGC 291  
 QY 224 GlyTyrLysAsnGlyTyrLysThrSerAlaGlyLysGluValIleAsnLysGluAsp 243  
 Db 292 GGTGGAGATTGCGTGGTGGAAAAAGCAGATGCAAGCTGTCTTAAATCAAGACCTA 351  
 QY 244 PheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGly 263  
 Db 352 TGGCAACGATTGCATGAGTCAACCAAAATCGCATTCATTGATGGCAATGGATCAAGGC 411  
 QY 264 HisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280  
 Db 412 CATGCTGGTCACTGGCAATGAGATGGCAGACCAACTGGCCCAATAAAGGC 462

## RESULT 2

US-09-596-002-40/c  
 ; Sequence 40, Application US/09596002  
 ; Patent No. 6632636  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lagace, Robert, E.  
 ; APPLICANT: Patterson, Chandra  
 ; APPLICANT: Berg, Kim, L.  
 ; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME  
 ; FILE REFERENCE: PM-0008-4 US  
 ; CURRENT APPLICATION NUMBER: US/09/596,002  
 ; CURRENT FILING DATE: 2000-06-16  
 ; PRIOR APPLICATION NUMBER: 60/140,121  
 ; PRIOR FILING DATE: 1999-06-18  
 ; NUMBER OF SEQ ID NOS: 41  
 ; SOFTWARE: PERL Program  
 ; SEQ ID NO 40  
 ; LENGTH: 119211  
 ; TYPE: DNA  
 ; ORGANISM: M. catarrhalis  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; OTHER INFORMATION: Incyte template ID No. 6632636 40  
 ; PUBLICATION INFORMATION:  
 US-09-596-002-40

Alignment Scores:  
 Pred. No.: 1,24e-17 Length: 119211  
 Score: 272.00 Matches: 62  
 Percent Similarity: 52.23% Conservative: 20  
 Best Local Similarity: 39.49% Mismatches: 51  
 Query Match: 17.59% Indels: 24  
 DB: 4 Gaps: 5

US-10-054-313-1 (1-286) x US-09-596-002-40 (1-119211)

QY 136 MetGlyAspPheValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLys 155  
 Db 45182 ATGAGCCAACTTAATATCGTTATACCGAGCGCGCTGTAAAGGCAATGGAACAAAGGC 45123  
 QY 156 ProArgAla---GlyIleGlyValTyr----- 163  
 Db 45122 GTATCTGCAGCGGTGGGCGGTGTTATTCATATTTCATGGCGATGAGCGGCATCTG 45063  
 QY 164 TrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGlnThrAsnGln 183  
 Db 45062 TGGGGC---GGTGAGCCT-----GATACGACCAATAAT 45033  
 QY 184 ArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsn 203

Db 45032 CGCATGGAATTGATGGCTGCCATCAGAGCTTTGGAGGCAACGCTGCACAG-----ATT 44979  
 QY 204 LysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGln 223  
 Db 44978 CCTCTGCACTTTGACAGATTGAGCTATGTTAAAGATGGCATAACTCAGTGGATTGGC 44919  
 QY 224 GlyTyrLysAsnGlyTyrLysThrSerAlaGlyLysGluValIleAsnLysGluAsp 243  
 Db 44918 GGTGGAGATTGCGTGGTGGAAAAAGCAGATGCAAGCTGTCTTAAATCAAGACCTA 44859  
 QY 244 PheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGly 263  
 Db 44858 TGGCAACGATTGCATGAGTCAACCAAAATCGCATTCATTGATGGCAATGGATCAAGGC 44799  
 QY 264 HisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280  
 Db 44798 CATGCTGGTCACTGGCAATGAGATGGCAGACCAACTGGCCCAATAAAGGC 44748

## RESULT 3

US-09-621-976-14380  
 ; Sequence 14380, Application US/09621976  
 ; Patent No. 6639063  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dumas Milne Edwards, J.B.  
 ; APPLICANT: Jobert, S.  
 ; APPLICANT: Giordano, J.Y.  
 ; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
 ; FILE REFERENCE: GENSET.054PR2  
 ; CURRENT APPLICATION NUMBER: US/09/621,976  
 ; CURRENT FILING DATE: 2000-07-21  
 ; NUMBER OF SEQ ID NOS: 19335  
 ; SOFTWARE: Patent.pm  
 ; SEQ ID NO 14380  
 ; LENGTH: 231  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; ORGANISM: Homo sapiens  
 US-09-621-976-14380

Alignment Scores:  
 Pred. No.: 2,41e-21 Length: 231  
 Score: 268.00 Matches: 63  
 Percent Similarity: 94.03% Conservative: 0  
 Best Local Similarity: 94.03% Mismatches: 4  
 Query Match: 17.34% Indels: 3  
 DB: 4 Gaps: 0

US-10-054-313-1 (1-286) x US-09-621-976-14380 (1-231)

QY 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg 20  
 Db 33 ATGAGCTGGCTTCGTTCCTGGCCCAAGAGTGGCTTGGCCGCTTGCCTGGCCGCCG 92  
 QY 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyArgLysThrGlyValPhe 40  
 Db 93 GGCTCTCGCG-ATCGGGATGTTCTATGCGCTGAG-AGGGCGCGCAAGACCGGGTCTTT 150  
 QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60  
 Db 151 CTGACCTGGATGATGATGAGAGCMAA-GTGAGCCGGTTCTTACTGTCAGATTAAAG 209  
 QY 61 PheAlaThrGluAspGluAla 67  
 Db 210 TTTGCCACAGAGGATGAGGC 230

## RESULT 4

US-09-252-991A-13638  
 ; Sequence 13638, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

```

; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13638
; LENGTH: 1275
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13638

Alignment Scores:
Pred. No.: 3,389-17 Length: 1275
Score: 241.00 Matches: 92
Percent Similarity: 35.96% Conservativity: 36
Best Local Similarity: 25.84% Mismatches: 126
Query Match: 15.59% Indels: 102
DB: 4 Gaps: 14

US-10-054-313-1 (1-286) x US-09-252-991A-13638 (1-1275)
QY 9 HisArgValAlaLeuAlaAlaLeuProCysArgArgGlySerArgGlyPheGlyWetPhe 28
DB 259 CACCGGCGAGATTACAGCGGG-----TGTCGGCTGGGGCGCGCTCGCGGGCGTGA 312
QY 29 TyrAlaValArgArgGlyValGlyThrGlyValPheLeuThrTrpAsnGluCysArgAla 48
DB 313 CATCCGCTCGGAAGAGGGCGC-----CTGGCGGCTCAGCGAACATCGCGCGGA 360
QY 49 GlnValAspArgPheProAlaAlaArg-----PheLysLysPheAlaThrGluAsp 65
DB 361 -----CGTGTCTCTGCTGCAACAGCGCTGGATTCTGCTGTGCTCTACCGTCT 411
QY 66 GluAlaTrpAla---PheValArgLysSerLaserProGluValSerGluGlyHisGlu 84
DB 412 COTGCGGAAGCGCGCGTACGTTCTGTCGGGGCGGCCACCTGCTGCTGATCGGCATCAA 471
QY 85 AsnGlnHis-----GlyGln 89
DB 472 CCCATGGAGCCTGTGGGCGATCCGTCATATTTCGCCGGGATGCTTGCGCCAGGCCG 531
QY 90 GluSerGluAlaLysProGlyLysArgLeuArgGluProLeuAspGlyAspGlyHisGlu 109
DB 532 CTGCAATCTCGTCCGCGCGCTGCGACTGGCTCAACCTGCTGGGCTTCGGCGTGGAGAA 591
QY 110 SerAla-----111
DB 592 ACGGCGCTTCGGGTGCTATGCTCGCGCGCTGTCGTCGCGCAGCTGGCAATCGCGCTG 651
QY 112 GlnProTyrAlaLysHisMetLys-----PheSerTyrMetGly---137
DB 652 TCGCCTGGAGCGCTGGGGCGAGCGCTGCGAGTCTTCGGGCGCGCTTCTATCTATTGTT 711
QY 120 -----ProSerValGluProAlaProProValSerArgAspThr-----132
DB 712 GGCACGCAAGCTGCTGCTGGGGTTCGCCCGCTGCGCCAGAGCAAGCGCGCAACCGCG 771
QY 133 -----PheSerTyrMetGly---137
DB 772 TCAGCTGTGGCCATCGCGGTGGCGAAGATCAGCGCGCGAGATTCGGAATTAGCATG 831
QY 138 -----AspPheValValValTyrThrAspGlyCysSerSerAsnGlyArgArg 154
DB 832 ACAGATAAAGACAGGTAGTATCTATACCGACGCGCGCTGCAAGGGCAACCTCGGGGCG 891
QY 155 LysProArgAlaGlyIleGlyValTyr-----TyrGlyProGly 167
DB 892 GCGGCTGGGGCGGTGTCTCTACAGGCGCGGAGCGAGCTTTGGGCG---GCG 948
QY 168 HisProLeuAsnValGlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGluIle 187

```

```

DB 949 GAGCCG-----GACACCACCAACCCGATGGAACTG 981
QY 188 HisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeu 207
DB 982 ATGCGCGGATCCAGCGGCTGGCGGCACCTCAAGCGTTCCTGCGGATCGTCTGATC---1038
QY 208 TyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLys 227
DB 1039 ---ACCGACTCGGAATACGTGATCGCGGCATCACCGAATGGTTGCCGAACCTGGAAGAAG 1095
QY 228 AsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeu 247
DB 1096 CGCGGCTGGAAGACCGCCAGCAGCGCGGTCAAGATGCCGACCTCTGGCAGGCGCTG 1155
QY 248 GluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPhe 267
DB 1156 GATGAGCAGGTCCCGCGGCAGGAGTGAGTGCGATGGGTCCGCGGGGATACCGGCGAT 1215
QY 268 IleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyAlaLysGln 283
DB 1216 CCCGGCAACGAGCGGCGCCACCGATTGGCCAAACCGTGGCGTGGCCGAA 1263

RESULT 5
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB186P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Alignment Scores:
Pred. No.: 7,4e-12 Length: 1830121
Score: 236.50 Matches: 54
Percent Similarity: 49.68% Conservativity: 23
Best Local Similarity: 34.84% Mismatches: 49

```



```

Query Match: 15.30% Indels: 29
DB: 4 Gaps: 4
US-10-054-313-1 (1-286) x US-09-557-884-1 (1-1830121)
QY 140 ValValValValThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGly 159
DB 152742 ATTGAATTTTACTGATGATCTTCTAGGTAAT-----CCAGGGCGGGC 152789
QY 160 IleGlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArg 179
DB 152790 -----GGAATTGGTGGCGGTATTGGTTATATA 152816
QY 180 Gln-----ThrAsnGlnArgAlaGluLeu 187
DB 152817 CAACATGAAAAACACCTCTCCAAAGCTATTTCACCAACCAATAATCGATGAATTA 152876
QY 188 HisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeu 207
DB 152877 CGCGCTGTCTATGAAGCATTAATACATTAAGAACACCTTGTCTG-----ATCAGCCTT 152930
QY 208 TyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLys 227
DB 152931 TATAGTGTAGCCATATATGAAAAATGCAATACCAATGATCTTTAACTGGAAAAA 152990
QY 228 AsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeu 247
DB 152991 AATAATTGAAAGCAAGTCTCGAAAGCCTGTAAAAAACCAAGATTATGGATAGCCTTA 153050
QY 248 GluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPhe 267
DB 153051 GATGAATCCATCCAACTCATATAAATTAATTTGCAATGGTAAAGGGCGACAC 153110
QY 268 IleGlyAsnGluAlaAspArgLeuAlaArgGluGlyAlaLys 282
DB 153111 AGAGAAAATGAATTTGCGATGAATTAGCAAAAAGGGCGAGAA 153155

RESULT 6
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21
ATTORNEY/AGENT INFORMATION:

```

```

NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PE186P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1
Alignment Scores:
Pred. NO.: 7.4e-12 Length: 1830121
Score: 236.50 Matches: 54
Percent Similarity: 49.68% Conservative: 23
Best Local Similarity: 34.84% Mismatches: 49
Query Match: 15.30% Indels: 29
DB: 4 Gaps: 4
US-10-054-313-1 (1-286) x US-09-643-990A-1 (1-1830121)
QY 140 ValValValValThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGly 159
DB 152742 ATTGAATTTTACTGATGATCTTCTAGGTAAT-----CCAGGGCGGGC 152789
QY 160 IleGlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArg 179
DB 152790 -----GGAATTGGTGGCGGTATTGGTTATATA 152816
QY 180 Gln-----ThrAsnGlnArgAlaGluLeu 187
DB 152817 CAACATGAAAAACACCTCTCCAAAGCTATTTCACCAACCAATAATCGATGAATTA 152876
QY 188 HisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeu 207
DB 152877 CGCGCTGTCTATGAAGCATTAATACATTAAGAACACCTTGTCTG-----ATCAGCCTT 152930
QY 208 TyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLys 227
DB 152931 TATAGTGTAGCCATATATGAAAAATGCAATACCAATGATCTTTAACTGGAAAAA 152990
QY 228 AsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeu 247
DB 152991 AATAATTGAAAGCAAGTCTCGAAAGCCTGTAAAAAACCAAGATTATGGATAGCCTTA 153050
QY 248 GluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPhe 267
DB 153051 GATGAATCCATCCAACTCATATAAATTAATTTGCAATGGTAAAGGGCGACAC 153110
QY 268 IleGlyAsnGluAlaAspArgLeuAlaArgGluGlyAlaLys 282
DB 153111 AGAGAAAATGAATTTGCGATGAATTAGCAAAAAGGGCGAGAA 153155

RESULT 7
US-09-639-637A-154
; Sequence 154, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Wiklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsoergalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361002

```

; CURRENT APPLICATION NUMBER: US/09/199,637A  
 ; CURRENT FILING DATE: 1998-11-25  
 ; PRIOR APPLICATION NUMBER: 60/066,517  
 ; PRIOR FILING DATE: 1997-11-25  
 ; NUMBER OF SEQ ID NOS: 437  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 154  
 ; LENGTH: 801  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-199-637A-154

Alignment Scores:  
 Pred. No.: 6,79e-17 Length: 801  
 Score: 235.50 Matches: 60  
 Percent Similarity: 49.12% Conservative: 24  
 Best Local Similarity: 35.09% Mismatches: 64  
 Query Match: 15.23% Indels: 24  
 DB: 4 Gaps: 5

US-10-054-313-1 (1-286) x US-09-199-637A-154 (1-801)

```

Qy 124 ProAlaProValSerArgAspThrPheSerTyrMetGly-----AspPhe 139
    |||
Db 314 CCGGTGGCGAAAGTCAGCGCGG-AGATTCCGAAATTTAGGCATGACAGATAAAGAACAG 372
    |||
Qy 140 ValValValTyrThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGly 159
    |||
Db 373 GTAGTGATCTATACCGACGGCGCTGCAAGGCGAACCTCGGCGCGCGGTGGGGGCG 432
    |||
Qy 160 IleGlyValTyr-----TrpGlyProGlyHisProLeuAsnVal 172
    |||
Db 433 TTGCTCTCTACAAAGCGCGCGAGAGAGCTTTGGGGC-----GGCGAGCGG----- 480
    |||
Qy 173 GlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGluIleHisAlaAlaCysLys 192
    |||
Db 481 -----GACACCCCAACACCGATCGAGCTTGGGGC-----GGCGAGCGG----- 480
    |||
Qy 193 AlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMet 212
    |||
Db 523 GCGCTGGCGGCACTCAAGGTTCTCTCGATCCGTCGTGATC-----ACCGACTCGGAA 576
    |||
Qy 213 PheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLysAsnGlyTrpLysThr 232
    |||
Db 577 TACGTGATGCGCGGATCACCGAATGTTGCGCACTCGAAGAGCGCGGTGGAAGACC 636
    |||
Qy 233 SerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGln 252
    |||
Db 637 GCACGACAGCAGCTGTCAAGATGCCGACCTCTGCGAGGCCCTGGATGACAGTCCGC 696
    |||
Qy 253 GlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGlu 272
    |||
Db 697 CGGCACCAAGTGGAGTGGAGTGGGTCCGCGGCGCATACCGGCGACCCCGGCAACAGAGCGG 756
    |||
Qy 273 AlaAspArgLeuAlaArgGluGlyAlaLysGln 283
    |||
Db 757 GCGGACCAAGTGGCCAAACCGTGGCGTCCGCCGAA 789
    |||
  
```

#### RESULT 8

US-09-199-637A-136/c  
 ; Sequence 136, Application US/09199637A  
 ; Patent No. 6355411  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ausubel, Frederick  
 ; APPLICANT: Goodman, Howard M.  
 ; APPLICANT: Rahme, Laurence G.  
 ; APPLICANT: Mahajan-Miklos, Shalina  
 ; APPLICANT: Tan, Man-Wah  
 ; APPLICANT: Cao, Hui  
 ; APPLICANT: Drenkard, Eliana  
 ; APPLICANT: Tsongalis, John  
 ; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID  
 ; SEQUENCES AND USES THEREOF

; FILE REFERENCE: 00786/361002  
 ; CURRENT APPLICATION NUMBER: US/09/199,637A  
 ; CURRENT FILING DATE: 1998-11-25  
 ; PRIOR APPLICATION NUMBER: 60/066,517  
 ; PRIOR FILING DATE: 1997-11-25  
 ; NUMBER OF SEQ ID NOS: 437  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 136  
 ; LENGTH: 2048  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-199-637A-136

Alignment Scores:  
 Pred. No.: 2,86e-16 Length: 2048  
 Score: 235.50 Matches: 60  
 Percent Similarity: 49.12% Conservative: 24  
 Best Local Similarity: 35.09% Mismatches: 64  
 Query Match: 15.23% Indels: 24  
 DB: 4 Gaps: 5

US-10-054-313-1 (1-286) x US-09-199-637A-136 (1-2048)

```

Qy 124 ProAlaProValSerArgAspThrPheSerTyrMetGly-----AspPhe 139
    |||
Db 648 CCGGTGGCGAAAGTCAGCGCGG-AGATTCCGAAATTTAGGCATGACAGATAAAGAACAG 590
    |||
Qy 140 ValValValTyrThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGly 159
    |||
Db 589 GTAGTGATCTATACCGACGGCGCTGCAAGGCGAACCTCGGCGCGCGGTGGGGGCG 530
    |||
Qy 160 IleGlyValTyr-----TrpGlyProGlyHisProLeuAsnVal 172
    |||
Db 529 TTGCTCTCTACAAAGCGCGCGAGAGAGCTTTGGGGC-----GGCGAGCGG----- 482
    |||
Qy 173 GlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGluIleHisAlaAlaCysLys 192
    |||
Db 481 -----GACACCCCAACACCGATCGAGCTTGGGGC-----GGCGAGCGG----- 482
    |||
Qy 193 AlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMet 212
    |||
Db 439 GCGCTGGCGGCACTCAAGGTTCTCTCGATCCGTCGTGATC-----ACCGACTCGGAA 386
    |||
Qy 213 PheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLysAsnGlyTrpLysThr 232
    |||
Db 385 TACGTGATGCGCGGATCACCGAATGTTGCGCACTCGAAGAGCGCGGTGGAAGACC 326
    |||
Qy 233 SerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGln 252
    |||
Db 325 GCACGACAGCAGCTGTCAAGATGCCGACCTCTGCGAGGCCCTGGATGACAGTCCGC 266
    |||
Qy 253 GlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGlu 272
    |||
Db 265 CGGCACCAAGTGGAGTGGAGTGGGTCCGCGGCGCATACCGGCGACCCCGGCAACAGAGCGG 206
    |||
Qy 273 AlaAspArgLeuAlaArgGluGlyAlaLysGln 283
    |||
Db 205 GCGGACCAAGTGGCCAAACCGTGGCGTCCGCCGAA 173
    |||
  
```

#### RESULT 9

US-09-252-991A-13802/c  
 ; Sequence 13802, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; TITLE OF INVENTION: SEQUENCES AND USES THEREOF



Db 148 GTAGAAATTTTCCACGCGCTCTTCCTGGTATCA----- 186  
Qy 160 IleGlyValTyrTrpGlyProGly----- 167  
Db 187 -----GGACCTGGCGGCTACGCGCCATATGCGCTACCGCACGAA 231  
Qy 168 HisProLeuAsnValGlyLeuArgLeuProGlyArgGlnThrAsnGlnArgAlaGluLe 187  
Db 232 AAACCTTCAGCGCTGGTACCGCTG-----ACCACATAACCGCATGCACTG 282  
Qy 188 HisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeu 207  
Db 283 ATGGCGGAATTTGTCCTCGAAGCGTTAAAGAGCAT-----TGGGAAGTGGTCTC 336  
Qy 208 TyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLys 227  
Db 337 AGCACTGACACCGATATGTCGTCAGGGGATCCACCGATGATCCCACTGGAAGAAAG 396  
Qy 228 AsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeu 247  
Db 397 CGCGGCTGAAACGGCAGAGAAAAGCGGTGAAAATGTCGACCTCTGGCAGCGCTG 456  
Qy 248 GluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPhe 267  
Db 457 GACGCGCGCTGGCCACGATATAAATGGAATGGGAATGGGTAAGGGCCATGCGCGCAT 516  
Qy 268 IleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyAla 281  
Db 517 CCTGAAATGAACGCTGCGACGAGCTGGCGCGCGCGCGCC 558

## RESULT 12

US-09-543-681A-2007  
; Sequence 2007, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543.681A  
; PRIOR FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 2007  
; LENGTH: 489  
; TYPE: DNA  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-2007

Alignment Scores:  
Pred. No.: 1.18e-12 Length: 489  
Score: 194.50 Matches: 50  
Percent Similarity: 45.34% Conservative: 23  
Best Local Similarity: 31.06% Mismatches: 59  
Query Match: 12.58% Indels: 29  
Gaps: 4

US-10-054-313-1 (1-286) x US-09-543-681A-2007 (1-489)

Qy 134 SerTyrMetGlyAspPheValValValTyrThrAspGlyCysSerSerAsnGlyArg 153  
Db 4 GCCTTTATGCACAGCAGTGAATAATTCACCGATGTTTCATGTTTAGGCAACCCA--- 60  
Qy 154 ArgLysProArgAlaGlyIleGlyValTyrTrpGlyProGlyHisProLeuAsnValGly 173  
Db 61 -----GGTCTGCTG----- 78  
Qy 174 IleArgLeuProGlyArgGln-----Thr 181  
Db 79 GCAATTTTACGCTACCAACAGCATGAAACCCCTTAGTGAAGGTTTTTTTATGACCAACC 138  
Qy 182 AsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsn 201

Db 139 AATAACCGCATGGAACCTCTTCCTGCTATCGATAGCAATTAAGCGGTAAATTCCTC--- 195  
Qy 202 IleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrp 221  
Db 196 ---TGTAATAATTACACTGACTACGATAGCCAGTATGTCCAGACAGGGAATTACCAAGTGG 252  
Qy 222 ValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLys 241  
Db 253 ATACATAGTTGGAATAAAGCCCAATGCGGTAAGCAGATAAAGCCCTGTGCTGATGTT 312  
Qy 242 GluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisVal 261  
Db 313 GATTTATGAAGCGCTTCATATAAGCCATTGAGCGTCATGAAATGAAATGCGATTGGGTT 372  
Qy 262 ProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyAla 281  
Db 373 AAAGGTCATGCGAGCGCATGACGAAATGAACCTTGTGATGAATGCGCAAGCGCGCGC 432  
Qy 282 Lys 282  
Db 433 CAA 435

## RESULT 13

US-08-163-181-3  
; Sequence 3, Application US/08163181  
; Patent No. 5459055  
; GENERAL INFORMATION:  
; APPLICANT: Jendrisak, Jerome J.  
; APPLICANT: Dahl, Robert E.  
; TITLE OF INVENTION: THERMOSTABLE RIBONUCLEASE H  
; TITLE OF INVENTION: AND GENETIC CONSTRUCT  
; TITLE OF INVENTION: THEREFORE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 411 East Wisconsin Avenue  
; CITY: Milwaukee  
; STATE: Wisconsin  
; COUNTRY: U.S.A.  
; ZIP: 53202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version  
; SOFTWARE: #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/163,181  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/815,095  
; FILING DATE: 27-DEC-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Baker, Jean C.  
; REGISTRATION NUMBER: P-35,433  
; REFERENCE/DOCKET NUMBER: 31-307-9001-1  
; TELEPHONE: (414) 277-5709  
; TELEFAX: (414) 277-5774  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 501 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-163-181-3

Alignment Scores:  
Pred. No.: 2.36e-10 Length: 501

```

; REFERENCE/DOCKET NUMBER: 31-307-9001-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5709
; TELEFAX: (414) 277-5774
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 501 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-465-161-3

Alignment Scores:
Pred. No.: 2,36e-10 Length: 501
Score: 174.00 Matches: 51
Percent Similarity: 51.35% Conservative: 25
Best Local Similarity: 34.46% Mismatches: 59
Query Match: 11.25% Indels: 14
DB: Gaps: 5

US-10-054-313-1 (1-286) x US-08-465-161-3 (1-501)
QY 140 ValValValValThrAspGlyCysCysSerSerAsnGlyArgLysProArgAlaGly 159
Db 25 GTGGCCCTTTCACCGAGGGGCTGTCTGGGAACC-CGGGCCCGG-----GGG 74
QY 160 IleGlyValTyxTrpGlyPro-----GlyHisProLeuAsnValGlyIleArgLeu 176
Db 75 GTGGGC-----GGCCCTCTCTCGCTTCCACGCCCGCACGAGAAGCTCTCTCCGGGG 125
QY 177 ProGlyArgGln-ThrAsnGlnArgAlaGluIleHisAlaAlyAlaIleGlu 196
Db 126 AGAGGCTGCACCAACAACCGCATGGAGCTCAAGGGCGCCATAGAGGGCTTAAGGC 185
QY 196 nAlaLyThrGlnAsnIleAsnLysIleValLeuTyxThrasPserMetPhetHrIleas 216
Db 186 CCTCAAGGAGCCT-----TGCGAGGTGGACCTCTACCGACGACGACTACTCTCAAGAA 239
QY 216 nGlyIleThr---AsnTipValGlnGlyTpLysLysAsnGlyTyxThrasPserMetPhetHrIleas 235
Db 240 GGCCCTTCCCGAGGGCTGTGGTAGGCTGGCGAAAAGGGGTGGCGGACGGCGAGGG 299
QY 235 YLysGluValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrgInGlyMetAs 255
Db 300 CAAGCCCGTGAATAAACCGACCTCTGGAGGGCCCTCTCTCGCATATGCCCCCACC 359
QY 255 pIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspAr 275
Db 360 GTGGCGCTTCCACTTCGTGAAGGGGCACACGGGCCACCCCGGAGAACGACGCGGTGGACCG 419
QY 275 gLeuAlaArgGluGlyAlaLys 282
Db 420 GGAGGGCGAGCGCCAGGCCAG 441

RESULT 15
US-08-163-181-4
; Sequence 4, Application US/08163181
; Patent No. 5459055
; GENERAL INFORMATION:
; APPLICANT: Jendrisak, Jerome J.
; APPLICANT: Smith, Robert E.
; APPLICANT: Dahl, Gary A.
; TITLE OF INVENTION: THERMOSTABLE RIBONUCLEASE H
; TITLE OF INVENTION: AND GENETIC CONSTRUCT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: U.S.A.

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version
; SOFTWARE: #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/465,161
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/815,095
; FILING DATE: 27-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Baker, Jean C.
; REGISTRATION NUMBER: P-35,433

```

```

; REFERENCE/DOCKET NUMBER: 31-307-9001-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5709
; TELEFAX: (414) 277-5774
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 501 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-465-161-3

Alignment Scores:
Pred. No.: 2,36e-10 Length: 501
Score: 174.00 Matches: 51
Percent Similarity: 51.35% Conservative: 25
Best Local Similarity: 34.46% Mismatches: 59
Query Match: 11.25% Indels: 14
DB: Gaps: 5

US-10-054-313-1 (1-286) x US-08-465-161-3 (1-501)
QY 140 ValValValValThrAspGlyCysCysSerSerAsnGlyArgLysProArgAlaGly 159
Db 25 GTGGCCCTTTCACCGAGGGGCTGTCTGGGAACC-CGGGCCCGG-----GGG 74
QY 160 IleGlyValTyxTrpGlyPro-----GlyHisProLeuAsnValGlyIleArgLeu 176
Db 75 GTGGGC-----GGCCCTCTCTCGCTTCCACGCCCCACGAGAAGCTCTCTCCGGGG 125
QY 177 ProGlyArgGln-ThrAsnGlnArgAlaGluIleHisAlaAlyAlaIleGlu 196
Db 126 AGAGGCTGCACCAACAACCGCATGGAGCTCAAGGGCGCCATAGAGGGCTTAAGGC 185
QY 196 nAlaLyThrGlnAsnIleAsnLysLeuValLeuTyxThrasPserMetPhetHrIleas 216
Db 186 CCTCAAGGAGCCT-----TGCGAGGTGGACCTCTACCGACGACGACTACTCTCAAGAA 239
QY 216 nGlyIleThr---AsnTipValGlnGlyT-TpLysLysAsnGlyTpLysTh-SerAlaGl 235
Db 240 GGCCCTCACCGAGGGCTGGCTGGAGGCTGGCGAAAAGGGGCTGGCGGACGGCGAGG 299
QY 235 YLysGluValIleAsnLysGluAspPheValAlaLeuGluArgLeuthrGlnGlyMetAs 255
Db 300 CAAGCCCGTGAATAAACCGACCTCTGGAGGGCCCTCTCCCTCGCCATGGCCCCCACC 359
QY 255 pIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspAr 275
Db 360 GTGGCGCTTCCACTTCGTGAAGGGGCACACGGGCCACCCCGGAGAACGACGCGTGGAC 419
QY 275 gLeuAlaArgGluGlyAlaLys 282
Db 420 GGAGGGCGAGCGCCAGGCCAG 441

RESULT 15
US-08-163-181-4
; Sequence 4, Application US/08163181
; Patent No. 5459055
; GENERAL INFORMATION:
; APPLICANT: Jendrisak, Jerome J.
; APPLICANT: Smith, Robert E.
; APPLICANT: Dahl, Gary A.
; TITLE OF INVENTION: THERMOSTABLE RIBONUCLEASE H
; TITLE OF INVENTION: AND GENETIC CONSTRUCT
; TITLE OF INVENTION: THEREFORE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: U.S.A.

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version
; SOFTWARE: #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/465,161
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/815,095
; FILING DATE: 27-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Baker, Jean C.
; REGISTRATION NUMBER: P-35,433

```

Search completed: March 23, 2004, 00:22:28  
Job time : 711 secs

ZIP: 53202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version  
SOFTWARE: #1-25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/163,181  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/815,095  
FILING DATE: 27-DEC-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Baker, Jean C.  
REGISTRATION NUMBER: P-35,433  
REFERENCE/DOCKET NUMBER: 31-307-9001-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (414) 277-5709  
TELEFAX: (414) 277-5774  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 552 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: Other Nucleic Acid  
US-08-163-181-4

## Alignment Scores:

Pred. No.:	2,74e-10	Length:	552
Score:	174.00	Matches:	51
Percent Similarity:	51.35%	Conservative:	25
Best Local Similarity:	34.46%	Mismatches:	59
Query Match:	11.25%	Indels:	14
DB:	1	Gaps:	5

US-10-054-313-1 (1-286) x US-08-163-181-4 (1-552)

QY	140	ValValValTyrThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGly	159
DB	76	GTGGCCCTCTTCACGACGGGGCTGCTGGGAACCC-CGGGCCCGG-----GGG	125
QY	160	IleGlyValTyrTrpGlyPro-----GlyHisProLeuAsnValGlyIleArgLeu	176
DB	126	GTGGGC-----GGCCCTCTCTCCGCTTCCAGCCCAAGAGCTCTCTCCGGGG	176
QY	177	ProGlyArgGln-ThrAsnGlnArgAlaIleHisAlaCysLysAlaIleGluG1	196
DB	177	AGAGCCTGACCCACCAACCCGATGGAGCTCAAGCGGCCATAGAGGCCCTAAAGGC	236
QY	196	nAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAs	216
DB	237	CCTCAAGGAGCCT-----TGGAGGTGGACTCTACACGACGACCACTACCTCAAGAA	290
QY	216	nGlyIleThr---AsnTrpValGlnGlyTyrLysLysAsnGlyTyrLysThrSerAlaG1	235
DB	291	GGCCTTCCAGGAGGCTGGTGAAGCTGGCGAAAGGGGCTGGCGACGGCGGAGGG	350
QY	235	YlysGluValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAs	255
DB	351	CAAGCCCGTGAACACCGGACCTCTGGGAGGCCCTCTCTCGCCATGGCCCCCACC	410
QY	255	pIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluAlaAspAr	275
DB	411	GGTGGCTTCCACTTCGTGAAGGGGCACACGGGCCACCGGGGACGAAACGAGCGGTGGACCG	470
QY	275	gLeuAlaArgGluGlyAlaLys	282
DB	471	GGAGCGGAGCGGCCAGGCCAG	492

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 22, 2004, 22:36:31 ; Search time 292 Seconds  
(without alignments)  
3623.143 Million cell updates/sec

Title: US-10-054-313-1

Perfect score: 1546

Sequence: 1 MSWLFARHVALPCCR.....FIGNEADRLREGAKQSED 286

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -p2n model -DEV-xlh

-Q-/cn2\_1/USPTO\_spool/US10054313/runat\_22032004\_140417\_28126/app\_query\_fasta\_1.455  
-DE-published Applications NA -QWTFastcap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom2  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-NAXLEN=200000000 -USRP=US10054313 @CGN 1.1 164 @runat\_22032004\_140417\_28126  
-NCPUS=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:  
1: /cn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:  
2: /cn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:  
3: /cn2\_6/ptodata/2/pubpna/US05\_NEW\_PUB.seq:  
4: /cn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:  
5: /cn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
6: /cn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:  
10: /cn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:  
11: /cn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:  
12: /cn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
13: /cn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:  
14: /cn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:  
15: /cn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:  
16: /cn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
17: /cn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:  
18: /cn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
-----					

	1	1526	98.7	1807	14	US-10-198-846-12889
	2	1522	98.4	965	12	US-10-263-511-85
	3	1045.5	67.6	631	12	US-10-263-511-79
	4	1045.5	67.6	716	12	US-10-263-511-77
	5	979.5	63.4	586	12	US-10-263-511-81
	6	792	51.2	457	12	US-10-263-511-83
	7	549	35.5	310	9	US-09-796-692-9423
	8	549	35.5	310	14	US-10-040-862-9423
	9	549	35.5	310	15	US-10-057-475B-9423
	10	549	35.5	310	15	US-10-154-884B-9423
	11	541	35.0	310	9	US-09-796-692-7011
	12	541	35.0	310	14	US-10-040-862-7011
	13	541	35.0	310	15	US-10-057-475B-7011
	14	541	35.0	310	15	US-10-154-884B-7011
	15	468	30.3	436	10	US-09-818-993-28506
	16	450	29.1	1279	10	US-09-814-353-13582
	17	434	28.1	473	9	US-09-960-352-6980
	18	399	25.8	407	9	US-09-960-352-6980
	19	364	23.5	764	9	US-09-910-943-431
	20	338.5	21.9	910	15	US-10-369-493-36958
	21	308	19.9	821	15	US-10-369-493-28857
	22	276	17.9	979	15	US-10-369-493-28827
	23	249.5	16.1	1047	15	US-10-369-493-25601
	24	240	15.5	499	10	US-09-814-353-13698
	25	240	15.5	501	10	US-09-814-353-943
	26	240	15.5	501	10	US-09-814-353-7313
	27	236.5	15.3	462	14	US-10-260-877-71
	28	236.5	15.3	1830121	14	US-10-329-960-1
	29	236.5	15.3	1830121	15	US-10-329-670-1
	30	235.5	15.2	801	10	US-09-975-719-154
	31	235.5	15.2	2048	10	US-09-975-719-136
	32	193.5	12.5	9025608	14	US-10-156-761-1
	33	176.5	11.4	708	14	US-10-156-761-1127
	34	171.5	11.1	1545	12	US-10-425-114-34337
	35	158	10.2	672	15	US-10-369-493-26705
	36	154	10.0	832	9	US-09-970-597-2
	37	152	9.8	8323	9	US-09-970-597-2
	38	150	9.7	185548	14	US-10-175-523-62
	39	144	9.3	8655	14	US-10-175-523-85
	40	141.5	9.2	640681	9	US-09-790-988-1
	41	140.5	9.1	13049	9	US-09-231-235-1
	42	140.5	9.1	13049	9	US-09-797-518A-1
	43	140.5	9.1	13049	10	US-09-996-073-1
	44	140.5	9.1	13049	10	US-09-872-696A-1
	45	138	8.9	7984	12	US-10-359-120-162

#### ALIGNMENTS

#### RESULT 1

US-10-198-846-12889  
; Sequence 12889, Application US/10198846  
; Publication No. US2003009974A1

; GENERAL INFORMATION:

; APPLICANT: Lillie, James

; APPLICANT: Xu, Yongyao

; APPLICANT: Wang, Youzhen

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS

; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND

; TITLE OF INVENTION: THERAPY OF BREAST CANCER

; FILE REFERENCE: MEI-049

; CURRENT APPLICATION NUMBER: US/10198,846

; CURRENT FILING DATE: 2002-07-18

; PRIOR APPLICATION NUMBER: 60/306,220

; PRIOR FILING DATE: 2001-07-18

; NUMBER OF SEQ ID NOS: 14084

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 12889

; LENGTH: 1807

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

NAME/KEY: misc\_feature  
LOCATION: 1802, 1803, 1804, 1805, 1806, 1807  
OTHER INFORMATION: n = A,T,C or G  
US-10-198-846-12889

Alignment Scores:  
Pred. No.: 1.66e-170 Length: 1807  
Score: 1536.00 Matches: 282  
Percent Similarity: 99.30% Conservative: 2  
Best Local Similarity: 98.60% Mismatches: 2  
Query Match: 98.71% Indels: 0  
DB: 14 Gaps: 0

US-10-054-313-1 (1-286) x US-10-198-846-12889 (1-1807)

```
QY 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
D 112 ATGAGCTGGCTTCTGTTCCCTGGCCACAGAGTGCCTTGGCCCTTGGCCCTGCCCGCC 171
QY 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyValArgValPhe 40
D 172 GGCTCTCGGGTTCGGATGTTCTATGCCGTGAGGAGGGCGCGCAAGACCGGGTCTTT 231
QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLys 60
D 232 CTGACCTGGAATGAGTCAGACAGACAGGTGGACCGGTTTCTGCTGCCAGATTAAAG 291
QY 61 PheAlaThrGluAspGluAlaTyrAlaPheValArgLysSerAlaSerProGluValSer 80
D 292 TTGCCCACAGAGTAGGCTGGCCCTTTGTGAGGAATCTGCAAGCCCGGAGTTTCA 351
QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
D 352 GAAGGGCATGAAATCAATCAATGACAAAGATCGAGCGGAAAGCCAGACGACTCCGT 411
QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
D 412 GAGCCATGATGAGATGAGCATGAAAGCGAGCGCGGTATGCAAGACATGAAGCCG 471
QY 121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
D 472 AGCGTGGAGCGCGCCCTCCAGTTAGCAGACAGCGTTTTCCTACATGGGAGACTTCGTC 531
QY 141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyLe 160
D 532 GTCGTCTACACTGATGCTCTGCTCCAGTAATGGCGGTAGAAGCGCGCGAGCAGGAATC 591
QY 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyLeArgLeuProGlyArgGln 180
D 592 GCGTTTACTGGGGCCAGGCCATCCTTTAATGTAGGCATTAGACTTCTTGGCGGCGAG 651
QY 181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200
D 652 ACAACCAAGAGCGGAAATTCATGACGCTGCAAGCCATTGAAACAAGCAAAAGACTCAA 711
QY 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
D 712 AACATCAATAACTGGTTCGTATACAGACAGATGTTTACGATAATGTGTATACTAAC 771
QY 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240
D 772 TGGTTCAGGTTTGAAGAAAATGGGTGGAAGACAGTCCAGGGAAGAGGTGATCAAC 831
QY 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
D 832 AAAGAGGACTTTGGGACCTGGAGAGCTTACCAGGGGATGGACATTCAGTGGATGCAT 891
QY 261 ValProGlyHisSerGlyPheIleGlyAsnGluAlaAspArgLeuAlaArgGluGly 280
D 892 GTTCTGTCATTCGGGATTTATAGGCAATGAAGAGCTGACAGATTAGCCAGAGAGA 951
QY 281 AlaLysGlnSerGluAsp 286
```

Db 952 GCTAAACAATCGGAGAC 969  
RESULT 2  
US-10-262-511-85  
Sequence 85, Application US/10262511  
Publication No. US20040038223A1  
GENERAL INFORMATION:  
APPLICANT: Smithson, Glenda  
APPLICANT: Millet, Isabelle  
APPLICANT: Peyman, John A.  
APPLICANT: Kekuda, Ramesh  
APPLICANT: Ju, Jingfang  
APPLICANT: Li, Li  
APPLICANT: Guo, Xiaojia (Sasha)  
APPLICANT: Patturajan, Meera  
APPLICANT: Spytek, Kimberly A.  
APPLICANT: Edinger, Shlomit R.  
APPLICANT: Ellerman, Karen  
APPLICANT: Malyankar, Uriel M.  
APPLICANT: Ort, Tatiana  
APPLICANT: Gorman, Linda  
APPLICANT: Zerhusen, Bryan D.  
APPLICANT: Anderson, David W.  
APPLICANT: Zhong, Mei  
APPLICANT: Catterton, Elina  
APPLICANT: Ji, Weizhen  
APPLICANT: Miller, Charles E.  
APPLICANT: Rastelli, Luca  
APPLICANT: Stone, David J.  
APPLICANT: Pena, Carol E. A.  
APPLICANT: Shenoy, Suresh G.  
APPLICANT: Shinkets, Richard A.  
APPLICANT: Rothenberg, Mark E.  
APPLICANT: Leach, Martin D.  
APPLICANT: Agee, Michele L.  
APPLICANT: Berghs, Constance  
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME  
FILE REFERENCE: 21402-462C  
CURRENT APPLICATION NUMBER: US/10/262,511  
CURRENT FILING DATE: 2003-05-28  
PRIOR APPLICATION NUMBER: 60/326,483  
PRIOR FILING DATE: 2001-10-02  
PRIOR APPLICATION NUMBER: 60/373,815  
PRIOR FILING DATE: 2002-04-19  
PRIOR APPLICATION NUMBER: 60/327,917  
PRIOR FILING DATE: 2001-10-09  
PRIOR APPLICATION NUMBER: 60/381,642  
PRIOR FILING DATE: 2002-05-17  
PRIOR APPLICATION NUMBER: 60/328,029  
PRIOR FILING DATE: 2002-10-09  
PRIOR APPLICATION NUMBER: 60/381,038  
PRIOR FILING DATE: 2002-05-16  
PRIOR APPLICATION NUMBER: 60/328,056  
PRIOR FILING DATE: 2001-10-09  
PRIOR APPLICATION NUMBER: 60/373,260  
PRIOR FILING DATE: 2002-04-17  
PRIOR APPLICATION NUMBER: 60/373,826  
PRIOR FILING DATE: 2002-04-19  
PRIOR APPLICATION NUMBER: 60/327,435  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 439  
SOFTWARE: CuraseqList version 0.1  
SEQ ID NO 85  
LENGTH: 965  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (10)...(867)  
US-10-262-511-85

Alignment Scores:



```
Pred. No.: 2,12e-170 Length: 965
Score: 1522.00 Matches: 281
Percent Similarity: 98.95% Conservative: 2
Best Local Similarity: 98.25% Mismatches: 3
Query Match: 98.45% Indels: 0
DB: 12 Gaps: 0

US-10-054-313-1 (1-286) x US-10-262-511-85 (1-965)

QY 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
DB 10 ATGAGCTGTTCTGTTCTGCGCCACAGAGTCGCTTGGCGGCTTGCCTGCGCGCG 69
QY 21 GlySerArgGlyPheGlyMetPheTrpAlaValArgGlyArgLysThrGlyValPhe 40
DB 70 GCCTCTCGCGGTTGCGGATGTTCTATGCGGTGAGAGGGGCCCAAGACCGGGGTCTTT 129
QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
DB 130 CTGACCTGGAATGAGTCAGAGCACAGGTGACCGGTTTCCTGCTGCCAGATTTAAGAAG 189
QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
DB 190 TTGCGCCAGAGGATGAGGCTTGGGCTTTGTGAGGAAATCTGCAAGCCCGGAATTCA 249
QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
DB 250 GAAGGGCATGAAATCAACATGGACAGAAATCGGAGCGAAGCCAGCAAGCGACTCGGT 309
QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
DB 310 GAGCCACTGGATGAGATGAGCATGAAAGCCGACGCGGTATGCAAGCCGCAATGAAGCCG 369
QY 121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
DB 370 AGCGTGGAGCGCGGCTCCAGTTAGCAGACACGTTTCTTACATGGGAGACTTGGTC 429
QY 141 ValValTyrThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGlyLe 160
DB 430 GTCGTCTACACTGATGCTGTGCTCCAGTAATGGCGGTAGAGCGCGCGAGCAGGAATC 489
QY 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyLeuArgLeuProGlyValGln 180
DB 490 GCGGTTTACTGGGGCCAGGCGATCTTTAAATGTAGGCATAGACTTCTTGGCGGCGAG 549
QY 181 ThrAsnGlnArgAlaGluLeuHisAlaAlaCysLysAlaLeuGluGlnAlaLysThrGln 200
DB 550 ACMAACCAAGAGCGGAATTCATGACGCTGCAAGCCATTGAACCAAGCAAGCACTCAA 609
QY 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
DB 610 AACATCAATAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATTAACCTAAC 669
QY 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240
DB 670 TGGGTTCAAGTTGGAAGAAATGGGTGGAAGACAGTGCAGGAAAGAGGTGATCAAC 729
QY 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
DB 730 AAAGAGGACTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAGTGGATGCAT 789
QY 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaAspGluGly 280
DB 790 GTTCTCTGTCATTCGGGATTTATAGGCAATGAGAGCTGACAGTAGCCAGAGAGGA 849
QY 281 AlalysGlnSerGluAsp 286
DB 850 GCTAAACAATCGGAAGAC 867
```

## RESULT 3

```
US-10-262-511-79
; Sequence 79, Application US/10262511
; Publication No. US2004003823A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Smithson, Glenda
; APPLICANT: Millet, Isabelle
; APPLICANT: Peyman, John A.
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Ju, Jingfang
; APPLICANT: Li, Li
; APPLICANT: Guo, Xiaojia (Sasha)
; APPLICANT: Patturajan, Meera
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Ort, Tatiana
; APPLICANT: Gorman, Linda
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Anderson, David W.
; APPLICANT: Zhong, Mei
; APPLICANT: Catterton, Elina
; APPLICANT: Ji, Weizhen
; APPLICANT: Miller, Charles E.
; APPLICANT: Rastelli, Luca
; APPLICANT: Stone, David J.
; APPLICANT: Pena, Carol E. A.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Rothenberg, Mark E.
; APPLICANT: Leach, Martin D.
; APPLICANT: Agee, Michele L.
; APPLICANT: Bergis, Constance
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-462C
; CURRENT APPLICATION NUMBER: US/10/262,511
; CURRENT FILING DATE: 2003-05-28
; PRIOR APPLICATION NUMBER: 60/326,483
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/373,815
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: 60/327,917
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/381,642
; PRIOR FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 60/328,029
; PRIOR FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: 60/381,038
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/328,056
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/373,260
; PRIOR FILING DATE: 2002-04-17
; PRIOR APPLICATION NUMBER: 60/373,826
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: 60/327,435
; PRIOR FILING DATE: 2001-10-05
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 439
; SOFTWARE: Curaseqlist version 0.1
; SEQ ID NO 79
; LENGTH: 631
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)..(631)
; US-10-262-511-79

Alignment Scores:
Pred. No.: 3,2e-114 Length: 631
Score: 1045.50 Matches: 201
Percent Similarity: 70.63% Conservative: 1
Best Local Similarity: 70.28% Mismatches: 1
Query Match: 67.63% Indels: 83
DB: 12 Gaps: 1
```

```
US-10-054-313-1 (1-286) x US-10-262-511-79 (1-631)
QY 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db 14 ATGAGCTGGTTCTGTCTTCCGCCACACAGTCGCTTGGCGCCCTTGGCCCTTGGCCGCCGCGC 73
QY 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgGlyThrGlyValPhe 40
Db 74 GGCTCTCGGGGTTCCGGATGTTCTATGCCGTGAGGAGGGCGCGCAAGACCCGGGCTCTTT 133
QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspGlyPheProAlaAlaArgPheLysLys 60
Db 134 CTGACCTGGATGAGTGC----- 151
QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 151 ----- 151
QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db 151 ----- 151
QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db 151 ----- 151
QY 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
Db 152 -----AGACACACGTTTCTTACATGGAGACATTCGTC 184
QY 141 ValValTyrThrAspGlyCysSerSerAsnGlyValArgLysProArgAlaGlyTle 160
Db 185 GTCGTCTACATGATGCTGCTCTCCAGTAAATGGCGTAGAGCGCGGACGAGGAATC 244
QY 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyLysArgLeuProGlyArgGln 180
Db 245 GCGCTTACTGGGGCGGGGCCCATCTCTTAAATGTAGGCATTAGACTTCTCTGGGGCGGAG 304
QY 181 ThrAsnGluArgAlaGluLysHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200
Db 305 ACAAACCAAGACGGGAATTCATGAGCGCTGCNAAGCCATTGACACAAAGAACTCAA 364
QY 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db 365 AACATCAATAAATCGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC 424
QY 221 TrpValGlnGlyTrpLysLysAsnGlyTyrLysThrSerAlaGlyLysGluValIleAsn 240
Db 425 TGGGTTCAAGTTGGAGAGAAAATGGTGGAGACAAAGTCAGGGAAGAGGTGATCAAC 484
QY 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Db 485 AAAGAGGACTTTTGGCACTGGAGAGGCTTACCCAGGGGATGCACATTCAGTGGATGCAT 544
QY 261 ValProGlyHisSerGlyPheIleGlyAsnGluAlaAspArgLeuAlaArgGluGly 280
Db 545 GTTCCCTGGTCAATTCGGGATTTATAGGCAATGAAGAGCTGACAGATTAGCCAGAGAAGA 604
QY 281 AlaLysGlnSerGluAsp 286
Db 605 GCTAAACAATCGGAAGAC 622

RESULT 4
US-10-262-511-77
; Sequence 77, Application US/10262511
; Publication No. US20040038223A1
; GENERAL INFORMATION:
; APPLICANT: Smithson, Glenna
; APPLICANT: Millett, Isabelle
; APPLICANT: Peyman, John A.
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Gu, Jingfang

US-10-054-313-1 (1-286) x US-10-262-511-77 (1-716)
QY 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db 10 ATGAGCTGGTTCTGTCTTCCGCCACACAGTCGCTTGGCGCCCTTGGCCCTTGGCCGCCGCGC 69

; APPLICANT: Li, Li
; APPLICANT: Guo, Xiaojia (Sasha)
; APPLICANT: Patturajan, Meera
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Efinger, Shlomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Ort, Tatiana
; APPLICANT: Gorman, Linda
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Anderson, David W.
; APPLICANT: Zhong, Mei
; APPLICANT: Catterton, Elina
; APPLICANT: Ji, Weizhen
; APPLICANT: Miller, Charles E.
; APPLICANT: Rastelli, Luca
; APPLICANT: Stone, David J.
; APPLICANT: Pena, Carol E. A.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Rothenberg, Mark E.
; APPLICANT: Leach, Martin D.
; APPLICANT: Agee, Michele L.
; APPLICANT: Berghs, Constance
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-462C
; CURRENT APPLICATION NUMBER: US/10/262,511
; CURRENT FILING DATE: 2003-05-28
; PRIOR APPLICATION NUMBER: 60/326,483
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/373,815
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: 60/327,917
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/381,842
; PRIOR FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 60/328,029
; PRIOR FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: 60/381,038
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/328,056
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/373,260
; PRIOR FILING DATE: 2002-04-17
; PRIOR APPLICATION NUMBER: 60/373,826
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: 60/327,435
; PRIOR FILING DATE: 2001-10-05
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 439
; SOFTWARE: CuraSeqlist version 0.1
; SEQ ID NO 77
; LENGTH: 716
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10)..(618)
; US-10-262-511-77

Alignment Scores:
Pred. No.: 3,79e-114 Length: 716
Score: 1045.50 Matches: 201
Percent Similarity: 70.63% Conservative: 1
Best Local Similarity: 70.28% Mismatches: 1
Query Match: 67.63% Indels: 83
DB: 12 Gaps: 1

US-10-054-313-1 (1-286) x US-10-262-511-77 (1-716)
QY 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db 10 ATGAGCTGGTTCTGTCTTCCGCCACACAGTCGCTTGGCGCCCTTGGCCCTTGGCCGCCGCGC 69
```

QY 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyArgGlyThrGlyValPhe 40  
 Db 70 GGCCTCGCGGTTTCGGGATGTTCTATCCGTGAGAGGGCCGCAAGACCGGGGTCTTT 129  
 QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60  
 Db 130 CTGACCTGGAATGAGTGC----- 147  
 QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80  
 Db 147 ----- 147  
 QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100  
 Db 147 ----- 147  
 QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120  
 Db 147 ----- 147  
 QY 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140  
 Db 148 -----AGAGACACGTTTCTCATCGGAGACTTCGTC 180  
 QY 141 ValValTyrThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGlyLe 160  
 Db 181 GTCGTCACACTGATGGCTGCTCCAGTAATGGCGTAGAGCGCGCGAGCAATC 240  
 QY 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyLeaArgLeuProGlyArgGln 180  
 Db 241 GCGGTTTACTGGGGCGCGGCATCTTTAAATGTAGGCATTAGACTTCCTGGCGGCAG 300  
 QY 181 ThrAsnGlnArgAlaGluLeuHisAlaLysAlaLysAlaLysGlnAlaLysThrGln 200  
 Db 301 ACAACCAAGAGCGGAATTCATGCAGCCCTGCAAGCCATTGAACCAAGCAAGACTCAA 360  
 QY 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyLeaThrAsn 220  
 Db 361 AACATCAATAACTGGTTCTGTATACACAGCATATGTTTACGATAAATGGTATACCTAAC 420  
 QY 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240  
 Db 421 TGGGTTCAAGTTGGAGAAATATAGGCAATGAAGAAGCTGACAGTGGAGAAAGAGTGCATCAAC 480  
 QY 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260  
 Db 481 AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCAGGGGATGGACATTCAGTGGATGCAT 540  
 QY 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280  
 Db 541 GTTCCTGTCATTCCGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGGA 600  
 QY 281 AlaLysGlnSerGluAsp 286  
 Db 601 GCTAAACAATCGGAAGAC 618

## RESULT 5

US-10-262-511-81  
 ; Sequence 81, Application US/10262511  
 ; Publication No. US2004003823A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Smithson, Glennda  
 ; APPLICANT: Millet, Isabelle  
 ; APPLICANT: Peyman, John A.  
 ; APPLICANT: Kekuda, Ramesh  
 ; APPLICANT: Ju, Jingfang  
 ; APPLICANT: Li, Li  
 ; APPLICANT: Guo, Xiaojia (Sasha)  
 ; APPLICANT: Fatturajan, Meera  
 ; APPLICANT: Spytek, Kimberly A.  
 ; APPLICANT: Edinger, Shlomiy R.  
 ; APPLICANT: Ellerman, Karen

; APPLICANT: Malyankar, Uriel M.  
 ; APPLICANT: Ort, Tatiana  
 ; APPLICANT: Gorman, Linda  
 ; APPLICANT: Zernhusen, Bryan D.  
 ; APPLICANT: Anderson, David W.  
 ; APPLICANT: Zhong, Mei  
 ; APPLICANT: Catterton, Elina  
 ; APPLICANT: Ji, Weizhen  
 ; APPLICANT: Miller, Charles E.  
 ; APPLICANT: Rastelli, Luca  
 ; APPLICANT: Stone, David J.  
 ; APPLICANT: Pena, Carol E. A.  
 ; APPLICANT: Shenoy, Suresh G.  
 ; APPLICANT: Shimkets, Richard A.  
 ; APPLICANT: Rothenberg, Mark E.  
 ; APPLICANT: Leach, Martin D.  
 ; APPLICANT: Agee, Michele L.  
 ; APPLICANT: Barghs, Constance  
 ; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME  
 ; FILE REFERENCE: 21402-462C  
 ; CURRENT APPLICATION NUMBER: US/10/262,511  
 ; CURRENT FILING DATE: 2003-05-28  
 ; PRIOR APPLICATION NUMBER: 60/326,483  
 ; PRIOR FILING DATE: 2001-10-02  
 ; PRIOR APPLICATION NUMBER: 60/373,815  
 ; PRIOR FILING DATE: 2002-04-19  
 ; PRIOR APPLICATION NUMBER: 60/327,917  
 ; PRIOR FILING DATE: 2001-10-09  
 ; PRIOR APPLICATION NUMBER: 60/381,642  
 ; PRIOR FILING DATE: 2002-05-17  
 ; PRIOR APPLICATION NUMBER: 60/328,029  
 ; PRIOR FILING DATE: 2002-10-09  
 ; PRIOR APPLICATION NUMBER: 60/381,038  
 ; PRIOR FILING DATE: 2002-05-16  
 ; PRIOR APPLICATION NUMBER: 60/328,056  
 ; PRIOR FILING DATE: 2001-10-09  
 ; PRIOR APPLICATION NUMBER: 60/373,260  
 ; PRIOR FILING DATE: 2002-04-17  
 ; PRIOR APPLICATION NUMBER: 60/373,826  
 ; PRIOR FILING DATE: 2002-04-19  
 ; PRIOR APPLICATION NUMBER: 60/327,435  
 ; PRIOR FILING DATE: 2001-10-05  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 439  
 ; SOFTWARE: CuraSeqList version 0.1  
 ; SEQ ID NO 81  
 ; LENGTH: 586  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (2)...(586)  
 ; US-10-262-511-81

Alignment Scores:  
 Pred. No.: 1,89e-106 Length: 586  
 Score: 979.50 Matches: 188  
 Percent Similarity: 69.60% Conservative: 2  
 Best Local Similarity: 68.86% Mismatches: 0  
 Query Match: 63.36% Indels: 83  
 DB: 12 Gaps: 1

US-10-054-313-1 (1-286) x US-10-262-511-81 (1-586)

QY 14 AlaAlaLeuProCysArgArgGlySerArgGlyPheGlyMetPheTyrAlaValArgArg 33  
 Db 8 TCCGCTTCCCTTCGCGCGCGGCTCTCGCGGTTCCGGATGTTCTATGCCCTGAGGAGG 67  
 QY 34 GlyArgLysThrGlyValPheLeuThrTrpAsnGluCysArgAlaGlnValAspArgPhe 53  
 Db 68 GGCCGCACAGACCGGGGTCTTTCTGACCTGGAATGAGTGC----- 106  
 QY 54 ProAlaAlaArgPheLysLysPheAlaThrGluAspGluAlaTrpAlaPheValArgLys 73

```
Db 106 ----- 106
Qy 74 SerAlaSerProGluValSerGluGlyHisGluAsnGlnHisGlyGlnSerGluAla 93
Db 106 ----- 106
Qy 94 LysProGlyLysArgLeuArgGluProLeuAspGlyAspGlyHisGluSerAlaGlnPro 113
Db 106 ----- 106
Qy 114 TyrAlaLysHisMetLysProSerValGluProAlaProValSerArgAspThrPhe 133
Db 107 -----AGAGACAGCTTT 118
Qy 134 SerTyrMetGlyAspPheValValValTyrThrAspGlyCysCysSerSerAsnGlyArg 153
Db 119 TCCTACATGGAGACTTCGTCGTCGTACACTGATGCTGCTCCAGTAATGGCGCT 178
Qy 154 ArgLysProArgAlaGlyIleGlyValTyrTyrGlyProGlyHisProLeuAsnValGly 173
Db 179 AGAAGCGCGGAGCGAGATCGCGCTTTACTGGGGCGCGGCCATCTTTTAAATGTAGGC 238
Qy 174 IleArgLeuProGlyArgGlnThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAla 193
Db 239 ATTAGACTTCCTGGCGGCGAGACAAACCAAGAGCGGAATTTCATCGACCTGCCAAGCC 298
Qy 194 IleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMetPhe 213
Db 299 ATTGAACAAGCAAAAGACTCAAAACATCAATAACTGGTTCTGTATACAGACAGTATGTT 358
Qy 214 ThrIleAsnGlyIleThrAsnTyrValGlnGlyTyrLysLysAsnGlyTyrLysThrSer 233
Db 359 ACATATAATGTTATTAATACTACTGGTTCAAGTTGGAGAAAATAATGGTGGAGACAGT 418
Qy 234 AlaGlyLysGluValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGlnGly 253
Db 419 GCAGGGAAGAGGTGATCAACAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGG 478
Qy 254 MetAspIleGlnTyrMetHisValProGlyHisSerGlyPheIleGlyAsnGluAla 273
Db 479 ATGGACANTTCAGTGGATGCGATGTTCTGTTCATCTCGGATTTATAGCAATGAAGAAGCT 538
Qy 274 AspArgLeuAlaArgGluGlyAlaLysGlnSerGluAsp 286
Db 539 GACAGATTAGCCAGAGAGAGGCTAAACAATCGGAAGAC 577
```

## RESULT 6

US-10-262-511-83

; Sequence 83, Application US/10262511

; Publication No. US20040038223A1

; GENERAL INFORMATION:

```
; APPLICANT: Smithson, Glennda
; APPLICANT: Millet, Isabelle
; APPLICANT: Peyman, John A.
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Ju, Jingfang
; APPLICANT: Li, Li
; APPLICANT: Guo, Xiaojia (Sasha)
; APPLICANT: Patturajan, Meera
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Edinger, Shomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Gorman, Linda
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Anderson, David W.
; APPLICANT: Zhong, Mei
; APPLICANT: Catterton, Elina
; APPLICANT: Ji, Weizhen
; APPLICANT: Miller, Charles E.
; APPLICANT: Rastelli, Luca
```

```
; APPLICANT: Stone, David J.
; APPLICANT: Pena, Carol E. A.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Rothenberg, Mark E.
; APPLICANT: Leach, Martin D.
; APPLICANT: Agee, Michele L.
; APPLICANT: Berghs, Constance
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-462C
; CURRENT APPLICATION NUMBER: US/10/262,511
; CURRENT FILING DATE: 2003-05-28
; PRIOR APPLICATION NUMBER: 60/326,483
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/373,815
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: 60/327,917
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/381,642
; PRIOR FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 60/328,029
; PRIOR FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: 60/381,038
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/328,056
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/373,260
; PRIOR FILING DATE: 2002-04-17
; PRIOR APPLICATION NUMBER: 60/373,826
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: 60/327,435
; PRIOR FILING DATE: 2001-10-05
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 439
; SOFTWARE: Curasequid version 0.1
; SEQ ID NO 83
; LENGTH: 457
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)..(457)
US-10-262-511-83
```

## Alignment Scores:

Pred. No.:	2.17e-84	Length:	457
Score:	792.00	Matches:	145
Percent Similarity:	100.00%	Conservative:	1
Best local Similarity:	99.32%	Mismatches:	0
Query Match:	51.23%	Indels:	0
DB:	12	Gaps:	0

US-10-054-313-1 (1-286) x US-10-262-511-83 (1-457)

Qy	137	GlyAspPheValValValTyrThrAspGlyCysCysSerSerAsnGlyArgAtgLysPro	156
Db	11	GGAGACTTCGTCGTCGTCTACACTGATGGCTGCTGCCTCCAGTAATGGCGTAGAAGCCG	70
Qy	157	ArgAlaGlyIleGlyValTyrTyrGlyProGlyHisProLeuAsnValGlyIleArgLeu	176
Db	71	CGAGCAGGAATCGGCGTTTACTGGGGCGCGCCATCTTTAAATGTAGGCATTAGACTT	130
Qy	177	ProGlyArgGlnThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGln	196
Db	131	CCTGGCGGCGAGACAAACCAAGAGCGGAATTCATCAGCCTCCAAAGCCATTGAACAA	190
Qy	197	AlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsn	216
Db	191	GCAAGACTCAAAACATCAATAAATCTGTATACACACAGTATGTTACGATAAT	250
Qy	217	GlyIleThrAsnTyrValGlnGlyTyrLysLysAsnGlyTyrLysThrSerAlaGlyLys	236
Db	251	GGTATACTAACTGGGTTCAAGGTTGAAGAAAAATGGGTGAAGACAAAGTCGACGGAAA	310

QY 237 GluValIleAenLysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIle 256  
DB 311 GAGGTGATCAACAAGAGCACTTTGTGCACTGGAGAGCTTACCCAGGGGATGACATT 370  
QY 257 GlnTrpMetHisValProGlyHisSerGlyPheIleGlyAenGluGluAlaAspArgLeu 276  
DB 371 CAGTGGATGCATGTTCTCGTGTTCATTCGGGATTTATAGCAATGAAGAAGCTGACAGATTA 430  
QY 277 AlaArgGluGlyAlaLys 282  
DB 431 GCCAGAGAGGAGCTAAA 448

## RESULT 7

US-09-796-692-9423/c  
; Sequence 9423, Application US/09796692  
; Publication No. US20020198362A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY  
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES  
; FILE REFERENCE: 2077.001200  
; CURRENT APPLICATION NUMBER: US/09/796,692  
; CURRENT FILING DATE: 2001-03-01  
; PRIOR APPLICATION NUMBER: 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: 60/223,416  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: 60/223,378  
; PRIOR FILING DATE: 2000-08-07  
; NUMBER OF SEQ ID NOS: 9597  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 9423  
; LENGTH: 310  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-796-692-9423

Alignment Scores:  
Pred. No.: 7,678-56 Length: 310  
Score: 549.00 Matches: 98  
Percent Similarity: 98.04% Conservative: 2  
Best Local Similarity: 98.04% Mismatches: 2  
Query Match: 35.51% Indels: 0  
DB: 9 Gaps: 0

US-10-054-313-1 (1-286) x US-09-796-692-9423 (1-310)

QY 68 TrpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHisGluAenGlnHis 87  
DB 308 TGGGCTTTGTTCAGGAATCTCAAGCCCGGAGTTTCAGAGGGCATGAAATCAACAT 249  
QY 88 GlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGluProLeuAspGlyAspGly 107

DB 248 GCACAAGATCGAGGCGAAAGCCAGCAAGCGACTCCGTGAGCACTGGATGGAGATGA 189  
QY 108 HisGluSerAlaGlnProTrpAlaLysHisMetLysProSerValGluProAlaProPro 127  
DB 188 CATGAAGCGCAGAGCCGTATGCAAGCACATGAAGCCGAGGTGGAGCCGCGCTCCA 129  
QY 128 ValSerArgAspThrPheSerTyrMetGlyAspPheValValValTyrThrAspGlyCys 147  
DB 128 GTTAGCAGACACAGCTTTCTTACATGGAGACTTCGTCGTCTACACTGATGGCTGC 69  
QY 148 CysSerSerAsnGlyArgArgLysProArgAlaGlyIleGlyValTyrTrpGlyProGly 167  
DB 68 TGCTCCAGTAATGGCGTAGAAGCGCGAGCAGGAATCGGCGTTTACTGGGCGCCAGGC 9  
QY 168 HisPro 169  
DB 8 CATCCT 3

## RESULT 8

US-10-040-862-9423/c  
; Sequence 9423, Application US/10040862  
; Publication No. US20030078396A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Retter, Marc  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Thera  
; TITLE OF INVENTION: Hematological Malignancies  
; FILE REFERENCE: 014058-013520US  
; CURRENT APPLICATION NUMBER: US/10/040,862  
; CURRENT FILING DATE: 2001-11-06  
; PRIOR APPLICATION NUMBER: US 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: US 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: US 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: US 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: US 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: US 60/223,416  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: US 60/223,378  
; PRIOR FILING DATE: 2000-08-07  
; PRIOR APPLICATION NUMBER: US 09/796,692  
; NUMBER OF SEQ ID NOS: 10467  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 9423  
; LENGTH: 310  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-040-862-9423

Alignment Scores:  
Pred. No.: 7,678-56 Length: 310  
Score: 549.00 Matches: 98  
Percent Similarity: 98.04% Conservative: 2  
Best Local Similarity: 98.04% Mismatches: 2

Query Match: 35.51% Indels: 0  
DB: 14 Gaps: 0  
US-10-054-313-1 (1-286) x US-10-040-862-9423 (1-310)  
QY 68 TtpAlaPheValArgLysSerAlaSerProGluValSerProGluValSerGluGlyHisGluAsnGlnHis 87  
Db 308 TGGGCTTTGTTCAGGAAATCTCAAGCCCGGAAGTTTCAGAGGGCGCATGAAATCAACAT 249  
QY 88 GlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGluProLeuAspGlyVasgLy 107  
Db 248 GGACAGAGATCGAGGCGAAGCCAGCAAGCACTCCGTGAGCCACTGGATGGATGGA 189  
QY 108 HisGluSerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaProPro 127  
Db 188 CATGAAGCGCAGACCGGTATGCAAGCACATGAAGCGGAGCGTGGAGCGCGCTCCA 129  
QY 128 ValSerArgAspThrPheSerTyrMetGlyAspPheValValValTyrThrAspGlyCys 147  
Db 128 GTTAGCAGACACAGCTTTCTTACATGGAGACTTCGTCTGTCTACACTGATGGCTGC 69  
QY 148 CysSerSerAsnGlyArgLysProArgAlaGlyIleGlyValTyrTrpGlyProGly 167  
Db 68 TGTCCAGTAATGGCGCTAGAGCGCGGAGCAGGAATCGGCGTTTACTGGGGGCCAGGC 9  
QY 168 HisPro 169  
Db 8 CATCCT 3  
RESULT 9  
US-10-057-475B-9423/c  
; Sequence 9423, Application US/10057475B  
; Publication No. US20040002068A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Clapper, Jonathan David  
; APPLICANT: Wang, Aijun  
; APPLICANT: Ordenez, Nadia  
; APPLICANT: Carter, Lauren  
; APPLICANT: McNeill, Patricia Dianne  
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy  
; FILE REFERENCE: 014058-014402US  
; CURRENT APPLICATION NUMBER: US/10/057,475B  
; CURRENT FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: US 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: US 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 10979  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 9423  
; LENGTH: 310

; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-057-475B-9423  
Alignment Scores:  
Pred. No.: 7,67e-56 Length: 310  
Score: 549.00 Matches: 98  
Percent Similarity: 98.04% Conservative: 2  
Best Local Similarity: 96.08% Mismatches: 2  
Query Match: 35.51% Indels: 0  
DB: 15 Gaps: 0  
US-10-054-313-1 (1-286) x US-10-057-475B-9423 (1-310)  
QY 68 TtpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHisGluAsnGlnHis 87  
Db 308 TGGGCTTTGTTCAGGAAATCTCAAGCCCGGAAGTTTCAGAGGGCGCATGAAATCAACAT 249  
QY 88 GlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGluProLeuAspGlyVasgLy 107  
Db 248 GGACAGAGATCGAGGCGAAGCCAGCAAGCACTCCGTGAGCCACTGGATGGATGGA 189  
QY 108 HisGluSerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaProPro 127  
Db 188 CATGAAGCGCAGACCGGTATGCAAGCACATGAAGCGGAGCGTGGAGCGCGCTCCA 129  
QY 128 ValSerArgAspThrPheSerTyrMetGlyAspPheValValValTyrThrAspGlyCys 147  
Db 128 GTTAGCAGACACAGCTTTCTTACATGGAGACTTCGTCTGTCTACACTGATGGCTGC 69  
QY 148 CysSerSerAsnGlyArgLysProArgAlaGlyIleGlyValTyrTrpGlyProGly 167  
Db 68 TGTCCAGTAATGGCGCTAGAGCGCGGAGCAGGAATCGGCGTTTACTGGGGGCCAGGC 9  
QY 168 HisPro 169  
Db 8 CATCCT 3  
RESULT 10  
US-10-154-884B-9423/c  
; Sequence 9423, Application US/10154884B  
; Publication No. US20040005561A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther  
; FILE REFERENCE: 014058-013521US  
; CURRENT APPLICATION NUMBER: US/10/154,884B  
; CURRENT FILING DATE: 2002-05-23  
; PRIOR APPLICATION NUMBER: US 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: US 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: US 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: US 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: US 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 60/222,903  
; PRIOR FILING DATE: 2000-08-03

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 11290  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 9423  
; LENGTH: 310  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-154-884B-9423

Alignment Scores:  
Pred. No.: 7,678-56 Length: 310  
Score: 549.00 Matches: 98  
Percent Similarity: 98.04% Conservative: 2  
Best Local Similarity: 96.08% Mismatches: 2  
Query Match: 35.51% Indels: 0  
DB: 15 Gaps: 0

US-10-054-313-1 (1-286) x US-10-154-884B-9423 (1-310)

Qy 68 TTPAlaPheValArgLysSerAlaSerProGluValSerGluGlyHisGluAsnGlnHis 87  
Db 308 TGGGCGCTTTGTCCAGGAATCTCAGAACCCGGAAGTTTCAGAAAGGCGCATGAAATCAACAT 249  
Qy 88 GlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGluProLeuAspGlyAspGly 107  
Db 248 GGACAGAATCCGAGCGGAAGACCCAGACGACTCCGTGAGCCACTCGATGGAGATGGA 189  
Qy 108 HisGluSerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaProPro 127  
Db 188 CATGAAGCGCAGAGCGGTATGCAAAAGCACATGAAGCGAGCGTGGAGCGCGGCTCCA 129  
Qy 128 ValSerArgAspThrPheSerTyrMetGlyAspPheValValTyrThrAspGlyCys 147  
Db 128 GTTAGCAGAGACACGTTTCTTACATGGAGACTTCGTCGTCTACACATGATGGCTGC 69  
Qy 148 CysSerSerAsnGlyArgLysProArgAlaGlyIleGlyValTyrTyrGlyProGly 167  
Db 68 TGCTCCAGTAATGGCGGTAGAGCGCGGAGCAGGAATCGCGTTTACTGGGGGCCAGGC 9  
Qy 168 HisPro 169  
Db 8 CATCCT 3

RESULT 11  
US-09-796-692-7011/c  
; Sequence 7011, Application US/09796692  
; Publication No. US20020198362A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Mannion, Jane  
; APPLICANT: Retter, Marc  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY  
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES  
; FILE REFERENCE: 2077.001200  
; CURRENT APPLICATION NUMBER: US/09/796,692  
; CURRENT FILING DATE: 2001-03-01  
; PRIOR APPLICATION NUMBER: 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: 60/218,950

; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: 60/223,416  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: 60/223,378  
; PRIOR FILING DATE: 2000-08-07  
; NUMBER OF SEQ ID NOS: 9597  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 7011  
; LENGTH: 310  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-796-692-7011

Alignment Scores:  
Pred. No.: 6,796-55 Length: 310  
Score: 541.00 Matches: 97  
Percent Similarity: 97.06% Conservative: 2  
Best Local Similarity: 95.10% Mismatches: 3  
Query Match: 34.99% Indels: 0  
DB: 9 Gaps: 0

US-10-054-313-1 (1-286) x US-09-796-692-7011 (1-310)

Qy 68 TTPAlaPheValArgLysSerAlaSerProGluValSerGluGlyHisGluAsnGlnHis 87  
Db 308 TGGGCGCTTTGTCCAGGAATCTCAGAACCCGGAAGTTTCAGAAAGGCGCATGAAATCAACAT 249  
Qy 88 GlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGluProLeuAspGlyAspGly 107  
Db 248 GGACAGAATCCGAGCGGAAGACCCAGACGACTCCGTGAGCCACTCGATGGAGATGGA 189  
Qy 108 HisGluSerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaProPro 127  
Db 188 CATGAAGCGCAGAGCGGTATGCAAAAGCACATGAAGCGAGCGTGGAGCGCGGCTCCA 129  
Qy 128 ValSerArgAspThrPheSerTyrMetGlyAspPheValValTyrThrAspGlyCys 147  
Db 128 GTTAGCAGAGACACGTTTCTTACATGGAGACTTCGTCGTCTACACATGATGGCTGC 69  
Qy 148 CysSerSerAsnGlyArgLysProArgAlaGlyIleGlyValTyrTyrGlyProGly 167  
Db 68 TGCTCCAGTAATGGCGGTAGAGCGCGGAGCAGGAATCGCGTTTACTGGGGGCCAGGC 9  
Qy 168 HisPro 169  
Db 8 CATCCT 3

RESULT 12

US-10-040-862-7011/c  
; Sequence 7011, Application US/10040862  
; Publication No. US20030078396A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Retter, Marc  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Thera  
; TITLE OF INVENTION: Hematological Malignancies  
; FILE REFERENCE: 014058-013520US  
; CURRENT APPLICATION NUMBER: US/10/040,862  
; CURRENT FILING DATE: 2001-11-06  
; PRIOR APPLICATION NUMBER: US 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: US 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,779

;; PRIOR FILING DATE: 2000-04-28  
;; PRIOR APPLICATION NUMBER: US 60/200,999  
;; PRIOR FILING DATE: 2000-05-01  
;; PRIOR APPLICATION NUMBER: US 60/202,084  
;; PRIOR FILING DATE: 2000-05-04  
;; PRIOR APPLICATION NUMBER: US 60/206,201  
;; PRIOR FILING DATE: 2000-05-22  
;; PRIOR APPLICATION NUMBER: US 60/218,950  
;; PRIOR FILING DATE: 2000-07-14  
;; PRIOR APPLICATION NUMBER: US 60/222,903  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: US 60/223,416  
;; PRIOR FILING DATE: 2000-08-04  
;; PRIOR APPLICATION NUMBER: US 60/223,378  
;; PRIOR FILING DATE: 2000-08-07  
;; PRIOR APPLICATION NUMBER: US 09/796,692  
;; PRIOR FILING DATE: 2001-03-01  
;; NUMBER OF SEQ ID NOS: 10467  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 7011  
;; LENGTH: 310  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-10-040-862-7011

Alignment Scores:  
Pred. No.: 6,79e-55 Length: 310  
Score: 541.00 Matches: 97  
Percent Similarity: 97.06% Conservative: 2  
Best Local Similarity: 95.10% Mismatches: 3  
Query Match: 34.99% Indels: 0  
DB: 14 Gaps: 0

US-10-054-313-1 (1-286) x US-10-040-862-7011 (1-310)

QY 68 TtpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHisGluAsnGlnHis 87  
Db 308 TGGGCTTTGTCAGGAATCTGCAAGCCGGAAGTTTCAGAGGGCATGAAATCAACAT 249  
QY 88 GlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGluProLeuAspGlyAspGly 107  
Db 248 GGACAAGAATCGGAGCGGAAGCCAGCAAGCGACTCCGTGAGCCACTGGATGGAGATGGA 189  
QY 108 HisGluSerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaProPro 127  
Db 188 CATGAAGCCAGACGCGATGTCAGACACATGAGTCGAGCGTGGAGCGGCGCTCCA 129  
QY 128 ValSerArgAspThrPheSerTyrMetGlyAspPheValValValTyrThrAspGlyCys 147  
Db 128 GTTAGCAGACACAGCTTTTCTCATGCGGAGACTTCGTGCTCTACACTGATGGCTGC 69  
QY 148 CysSerSerAsnGlyArgArgLysProArgAlaGlyIleGlyValTyrTrpGlyProGly 167  
Db 68 TGCTCCAGTATGGCGTAGAAGCCCGAGCAGAGATCGGCTTTTACTGGGGCGCGGC 9  
QY 168 HisPro 169  
Db 8 CATCCT 3

RESULT 13

US-10-057-475B-7011/c  
; Sequence 7011, Application US/10057475B  
; Publication No. US20040002069A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Clapper, Jonathan David  
; APPLICANT: Wang, AiJun  
; APPLICANT: Ordonez, Nadia  
; APPLICANT: Carter, Lauren  
; APPLICANT: McNeill, Patricia Dianne  
; APPLICANT: Corixa Corporation

;; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Thera  
;; TITLE OF INVENTION: Hematological Malignancies  
;; FILE REFERENCE: 014058-014402US  
;; CURRENT APPLICATION NUMBER: US/10/057,475B  
;; CURRENT FILING DATE: 2002-01-22  
;; PRIOR APPLICATION NUMBER: US 60/186,126  
;; PRIOR FILING DATE: 2000-03-01  
;; PRIOR APPLICATION NUMBER: US 60/190,479  
;; PRIOR FILING DATE: 2000-03-17  
;; PRIOR APPLICATION NUMBER: US 60/200,545  
;; PRIOR FILING DATE: 2000-04-27  
;; PRIOR APPLICATION NUMBER: US 60/200,303  
;; PRIOR FILING DATE: 2000-04-28  
;; PRIOR APPLICATION NUMBER: US 60/200,779  
;; PRIOR FILING DATE: 2000-04-28  
;; PRIOR APPLICATION NUMBER: US 60/200,999  
;; PRIOR FILING DATE: 2000-05-01  
;; PRIOR APPLICATION NUMBER: US 60/202,084  
;; PRIOR FILING DATE: 2000-05-04  
;; PRIOR APPLICATION NUMBER: US 60/206,201  
;; PRIOR FILING DATE: 2000-05-22  
;; PRIOR APPLICATION NUMBER: US 60/218,950  
;; PRIOR FILING DATE: 2000-07-14  
;; PRIOR APPLICATION NUMBER: US 60/222,903  
;; PRIOR FILING DATE: 2000-08-03  
;; Remaining Prior Application data removed - See File Wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 10979  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 7011  
;; LENGTH: 310  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-10-057-475B-7011

Alignment Scores:

Pred. No.: 6,79e-55 Length: 310  
Score: 541.00 Matches: 97  
Percent Similarity: 97.06% Conservative: 2  
Best Local Similarity: 95.10% Mismatches: 3  
Query Match: 34.99% Indels: 0  
DB: 15 Gaps: 0

US-10-054-313-1 (1-286) x US-10-057-475B-7011 (1-310)

QY 68 TtpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHisGluAsnGlnHis 87  
Db 308 TGGGCTTTGTCAGGAATCTGCAAGCCGGAAGTTTCAGAGGGCATGAAATCAACAT 249  
QY 88 GlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGluProLeuAspGlyAspGly 107  
Db 248 GGACAAGAATCGGAGCGGAAGCCAGCAAGCGACTCCGTGAGCCACTGGATGGAGATGGA 189  
QY 108 HisGluSerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaProPro 127  
Db 188 CATGAAGCCAGACGCGTATGCAAGACACATGAGTCGAGCGTGGAGCGGCGCTCCA 129  
QY 128 ValSerArgAspThrPheSerTyrMetGlyAspPheValValValTyrThrAspGlyCys 147  
Db 128 GTTAGCAGACACAGCTTTTCTCATGCGGAGACTTCGTGCTCTACACTGATGGCTGC 69  
QY 148 CysSerSerAsnGlyArgArgLysProArgAlaGlyIleGlyValTyrTrpGlyProGly 167  
Db 68 TGCTCCAGTATGGCGTAGAAGCCCGAGCAGAGATCGGCTTTTACTGGGGCGCGGC 9  
QY 168 HisPro 169  
Db 8 CATCCT 3

RESULT 14

US-10-154-884B-7011/c  
; Sequence 7011, Application US/10154884B  
; Publication No. US20040005561A1  
; GENERAL INFORMATION:



```

; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc W.
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-013521US
; CURRENT APPLICATION NUMBER: US/10/154.884B
; CURRENT FILING DATE: 2002-05-23
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 11290
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7011
; LENGTH: 310
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-154-884B-7011

Alignment Scores:
Pred. No.: 6,79e-55 Length: 310
Score: 541.00 Matches: 97
Percent Similarity: 97.06% Conservative: 2
Best Local Similarity: 95.10% Mismatches: 3
Query Match: 34.99% Indels: 0
DB: 15 Gaps: 0

US-10-054-313-1 (1-286) x US-10-154-884B-7011 (1-310)
QY 68 TrpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHisGluAsnGlnHis 87
Db 308 TGGGCTTTGTTCAGGAATCTGCAAGCCCGGAAGTTTCAGAAGGCGCATGAAATCAACAT 249
QY 88 GlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGluProLeuAspGlyAspGly 107
Db 248 GGACAAGAATCGAGCGCGAAAGCCAGCAAGGACTCCGTGAGCCACTGGATGGAGATGGA 189
QY 108 HisGluSerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaProPro 127
Db 188 CATGAAGCGCAGAGCCGTATGCAAGCACATGAGTCGAGCGTGGAGCCGCGGCTCCA 129
QY 128 ValSerArgAspThrPheSerTyrMetGlyAspPheValValValTyrThrAspGlyCys 147
Db 128 GTTAGCAGAGACACGTTTTCTTACATGGAGACTTCGTCGTCGTCTACACTGATGGCTGC 69
QY 148 CysSerSerAsnGlyValArgLysProArgAlaGlyIleGlyValTyrTrpGlyProGly 167
Db 68 TGCTCAGTAATGGCGTAGAAGCGCGGAGCAGCAAGAAATCGGCGTTTACTGGGGCGCGGC 9
QY 168 HisPro 169
Db 8 CATCCT 3
```

```

RESULT 15
US-09-918-995-29506
; Sequence 29506, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 29506
; LENGTH: 436
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(436)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-29506

Alignment Scores:
Pred. No.: 4,74e-46 Length: 436
Score: 468.00 Matches: 86
Percent Similarity: 100.00% Conservative: 3
Best Local Similarity: 96.63% Mismatches: 0
Query Match: 30.27% Indels: 0
DB: 10 Gaps: 0

US-10-054-313-1 (1-286) x US-09-918-995-29506 (1-436)
QY 198 LysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGly 217
Db 32 CAGACTCAAAACATCAATAAAGTCTGTTCTGTATACAGACAGATGTTTACTTTAAATGCT 91
QY 218 IleThrAsnTrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGlu 237
Db 92 ATTCTTAAGTGGTTCAAGGTTGGAAGAAAATGGTGGAGAACAGTCCAGGGAAAGAG 151
QY 238 ValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGln 257
Db 152 GTGATCAACAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTGAG 211
QY 258 TrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAla 277
Db 212 TGGATGCGATGTTCTCTGTCATTTCGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCC 271
QY 278 ArgGluGlyAlaLysGlnSerGluAsp 286
Db 272 AGAGAAGAGCTTAACAATCGGNAGAC 298

Search completed: March 23, 2004, 00:17:10
Job time : 303 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 22, 2004, 22:30:37 ; Search time 2335 Seconds  
(without alignments)  
3657.638 Million cell updates/sec

Title: US-10-054-313-1  
Perfect score: 1546  
Sequence: 1 MSWLFLLHVRVALAALPCRR.....FIGNEADRLAREGAKOSD 286

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame, p2n, model -DEV=xlh  
-Q/cgn2\_1/USFTO\_spoos/US10054313/runat\_22032004\_140416\_28082/app\_query.fasta\_1.455  
-DB=EST -CPMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10054313 @CGN 1.1 2135 @runat\_22032004\_140416\_28082 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: em estba.\*  
2: em esthum.\*  
3: em estin.\*  
4: em estmu.\*  
5: em estov.\*  
6: em estpl.\*  
7: em estro.\*  
8: em htc.\*  
9: gb est1.\*  
10: gb est2.\*  
11: gb htc.\*  
12: gb est3.\*  
13: gb est4.\*  
14: gb est5.\*  
15: em estfun.\*  
16: em eastom.\*  
17: em gss hum.\*  
18: em gss inv.\*  
19: em gss pin.\*  
20: em gss vrt.\*  
21: em gss fun.\*  
22: em gss nam.\*  
23: em gss mus.\*  
24: em gss pro.\*  
25: em gss rod.\*  
26: em gss phg.\*  
27: em gss vrl.\*  
28: gb gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1501	97.1	1201	13	BX394121
2	1498	96.9	1201	9	AL560874
3	1458	94.1	1054	12	BM541420
4	1446	93.5	1060	12	BM510851
5	1432	92.6	952	9	AL554334
6	1415.5	91.6	944	12	BG829086
7	1406	90.9	1196	13	BX416456
8	1345	87.0	893	13	BUI75613
9	1340	86.7	949	12	BG831321
10	1288	83.3	765	14	CA775004
11	1281	82.9	935	13	BQ211568
12	1239.5	80.2	779	10	BE746490
13	1229	79.5	818	13	BU595053
14	1222	79.0	916	12	BM451915
15	1215	78.6	869	10	BF984400
16	1210	78.3	707	12	BI223765
17	1200	77.6	959	10	BF689462
18	1198.5	77.5	714	10	BE778327
19	1195	77.3	946	10	BF038506
20	1155	74.7	764	12	BG767598
21	1153	74.6	1198	12	BM805842
22	1110.5	71.8	693	12	BG768406
23	1106.5	71.6	828	29	AY404958
24	1100	71.2	864	12	BG420594
25	1090.5	70.5	782	13	BU929741
26	1089.5	70.5	772	14	CB961021
27	1083.5	70.1	792	10	BE407918
28	1067	69.0	613	12	BM782115
29	1058	68.4	652	12	BI518784
30	1058	68.4	686	12	BG824728
31	1052.5	68.1	937	10	BE786259
32	1040.5	67.3	704	10	BF129039
33	1018	65.8	1209	12	BM454170
34	1003.5	64.9	868	13	BQ942688
35	1002	64.8	783	14	CB529180
36	995	64.4	1364	11	AK011680
37	992	64.2	927	12	BI558492
38	973	62.9	823	13	BQ96738
39	968.5	62.6	1102	12	BI410112
40	967	62.5	619	12	BG719686
41	964	62.4	857	14	CF617505
42	960.5	62.1	718	12	BG778946
43	952	61.6	801	10	BF690400
44	949	61.4	534	12	BM839989
45	945	61.1	807	9	AW006811

# ALIGNMENTS

RESULT 1  
LOCUS BX394121  
DEFINITION BX394121 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens  
CDNA clone CS0DC010Y123 5-PRIME, mRNA sequence.  
ACCESSION BX394121  
VERSION BX394121.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1201)

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Published (2001)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 1105.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CSODC010AE12QP1&cluster=1105.r. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CSODC010AE12QP1.

FEATURES  
Location/Qualifiers  
1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODC010Y123"  
/tissue\_type="NEUROBLASTOMA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo (dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN  
Alignment Scores:  
Pred. No.: 1.67e-130 Length: 1201  
Score: 1501.00 Matches: 281  
Percent Similarity: 99.30% Conservative: 3  
Best Local Similarity: 98.25% Mismatches: 2  
Query Match: 97.09% Indels: 1  
DB: 13 Gaps: 0

US-10-054-313-1 (1-286) x BX394121 (1-1201)

QY 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20  
:::  
:::  
99 YTGAGTGGCTTGTGTTCTTGGCCACAGAGTCGCTTGGCCGCTTGGCCGCCGCG 158  
21 GlySerArgGlyPheGlyMetPheTyrrAlaValArgGlyArgGlyThrGlyValPhe 40  
159 GGCTCTCGCGGGTTCGGGATGTTCTATGCGCTGAGGAGGGCGCGAAGACCGGGGCTTT 218  
41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60  
219 CTGACCTGGATGAGTGCAGAGCACAGGTGGACCGGTTTCTGCTCCAGATTAGAAG 278  
61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80  
279 TTTCACACAGAGGATGAGCGTGGCCCTTTGTACAGAAATCTCAAGCCCGGAAGTTTCA 338  
81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100  
339 GAAGCGCATGAATCATACATGCACAGAGATCGGAGCGGAAGT-ACCAAGCGACTCCGT 397  
101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrrAlaLysHisMetLysPro 120  
398 GAGCCACTGGATGGAGATGGACATGAAGCGCAGACCGCGTATCAAGACACATGAAGCCG 457  
121 SerValGluProAlaProProValSerArgAspThrPheSerTyrrMetGlyAspPheVal 140  
458 AGCGTGGAGCGGCGCCTCCAGTTAGCAGAGACACGTTTTTCTACATGGGAGACTTCGTC 517  
141 ValValTyrrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160  
518 GTCGCTACACTGATGCTGCTGCTCAGTAATGGCGGTAGAAGCGCGCAGCAGGAATC 577  
161 GlyValTyrrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180

578 GGCGTTTACTGGGGCCAGCCCATCTTTAAATGTAGGCATTAGACTTCTCTGGCGCGCAG 631

181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysGlyValAlaIleGluGlnAlaLysThrGln 200

638 ACAAAACCAAGAGCGGAATTCATGCAGCCTGCAAGCCATTGAACAAGCAAGACTCAA 697

201 AsnIleAsnLysLeuValLeuTyrrThrAspSerMetPheThrIleAsnGlyIleThrAsn 240

698 AACATCAATAAATCTGGTTCTGTATACAGACAGATGTTTACGATAAAATGTTATAACTAAC 757

221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240

758 TGGGTTTCAGGTTGGAAGAAATGGGTGGAAGCAAGTCCAGGAAGAAGAGGTGATCAAC 817

241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260

818 AAGAGGACATTTGTGGCACTGGAGAGGCTTACCAGGGGATGCACATTCACTAGTGATGCAT 877

261 ValProGlyHisSerGlyPheIleGlyAsnGluAlaAspArgLeuAlaArgGluGly 280

878 GTTCTCTGGTCACTCGGATTTATAGGCATGAAGAGCTGCACAGATTAGCAGAGAAGGA 937

281 AlalysGlnSerGluAsp 286

938 GCTAAACAATCGGAAGAC 955

RESULT 2  
AL560874  
LOCUS  
DEFINITION  
AL560874 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
Homo sapiens cDNA clone CSODL005YJ14 5-PRIME, mRNA sequence.  
ACCESSION  
AL560874  
VERSION  
AL560874.2 GI:31285003  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (Bases 1 to 1201)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Published (2001)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 1105.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CSODL005DE07QP1&cluster=1105.r. Contact :  
Feng Liang Email : fliang@lifetech.com URL : Corporation 1600  
Faraday Avenue Genoscope sequence ID : CSODL005DE07QP1.

FEATURES  
Location/Qualifiers  
1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODL005YJ14"  
/cell\_type="B CELLS (RAMOS CELL LINE)"  
/cell\_line="RAMOS CELL LINE"  
/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo (dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN  
Alignment Scores:  
Pred. No.: 3.58e-130 Length: 1201

Score: 1498.00 Matches: 281  
 Percent Similarity: 98.95% Conservative: 2  
 Best Local Similarity: 98.25% Mismatches: 3  
 Query Match: 96.90% Indels: 1  
 DB: 9 Gaps: 0

US-10-054-313-1 (1-286) x AU560874 (1-1201)

QY 1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20  
 DB 142 ATGAGCTGGCTTCGTCTCTGGCCACAGAGTCGCTTGGCGGCTTGCCTTGGCGCGC 201  
 QY 21 GlySerArgGlyPheGlyMetPheTyraValArgArgGlyArgLysThrGlyValPhe 40  
 DB 202 GCCTCTCGCGGTTCTGGGATGTTCTATCGCTGAGAGCGGCGCGCAAGACCGGGGTCTTT 261  
 QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60  
 DB 262 CTGACTCTGGATGATGATGAGACAGACAGGTGACCGGTTCTGCTGCCAGATTTAGAG 321  
 QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80  
 DB 322 TTTGCCACAGAGGATGAGGCTTGGGCTTTGTTCAGGAAATCTGCAAGCGCGGAAGTTCA 381  
 QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100  
 DB 382 GAAGGCGATGAATCAACATGACAGAAATCGAGAGCGGAAGC-AGCAAGCGACTCCGT 440  
 QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyraAlaLysHisMetLysPro 120  
 DB 441 GAGCCACTGGATGGAGATGGACATGAAAGCGCAGAGCGGTATGCAAGCACATGAAGCGC 500  
 QY 121 SerValGluProAlaProValSerArgAspThrPheSerTyrrMetGlyAspPheVal 140  
 DB 501 AGCGTGGAGCGCGGCTCCAGTTAGCAGACACAGTTTCTCTACATGGGAGCTTGGTC 560  
 QY 141 ValValTyrrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160  
 DB 561 GTTCGTCTACATGATGGCTGCTCCAGTAATGGCGTAGAGCGCGGAGCGAATC 620  
 QY 161 GlyValTyrrGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180  
 DB 621 GCGGTTTACTGGGGGCCAGGCCATCTTTAAATGTAGGCATTAGACTTCTCTGGCGGCGAG 680  
 QY 181 ThrAsnGlnArgAlaGluIleHisAlaLaCysLysAlaIleGluGlnAlaLysThrGln 200  
 DB 681 ACAACCAAGAGCGGAATTCATGCGCTCCAAAGCCATTCACAGCAAGACTCA 740  
 QY 201 AsnIleAsnLysLeuValLeuTyrrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220  
 DB 741 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTAACTAAC 800  
 QY 221 TrpValGlnGlyTrpLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240  
 DB 801 TGGGTTCAAGGTTGGAAGAAATGGTGGAGAGCAAGTGCAGGGAAGAGTGATCAAC 860  
 QY 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260  
 DB 861 AAAGAGCACTTGTGTCATGAGAGGCTTACCCAGGGGATGACATTCAGTGGATGCAT 920  
 QY 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280  
 DB 921 GTTCTGTGTCATTCGGGATTTATAGCAATGAGAAGCTGACAKATTAGCCAGAGAGGA 980  
 QY 281 AlaLysGlnSerGluAsp 286  
 DB 981 GCTAAACAATCGGAGAC 998

RESULT 3

BM541420

LOCUS

DEFINITION

BM541420 1054 bp mRNA linear EST 20-FEB-2002  
 AGENCOURT 6493820 NIH\_MGC\_71 Homo sapiens cDNA clone IMAGE:5521285  
 5', mRNA sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

BM541420  
 BM541420.1 GI:18770049  
 EST.  
 Homo sapiens (human)

Homo sapiens

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-femail.nih.gov

Tissue Procurement: ARCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM12186 row: k column: 14

High quality sequence stop: 713.

Location/Qualifiers

source

1. .1054

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:5521285"

/tissue\_type="leiomyosarcoma"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_71"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;

Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 2.1 kb.

ORIGIN

Alignment Scores:

Pred. No.: 3,24e-126 Length: 1054

Score: 1455.00 Matches: 278

Percent Similarity: 97.24% Conservative: 4

Best Local Similarity: 95.86% Mismatches: 4

Query Match: 94.11% Indels: 4

DB: 12 Gaps: 0

US-10-054-313-1 (1-286) x BM541420 (1-1054)

QY 1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20  
 DB 45 ATGAGCTGGCTTCGTCTCTGGCCACAGAGTCGCTTGGCGGCTTGCCTTGGCGCGC 104  
 QY 21 GlySerArgGlyPheGlyMetPheTyraValArgArgGlyArgLysThrGlyValPhe 40  
 DB 105 GGCTCTCGCGGTTCCGGATGTTCTATGCGTGAGAGGGCGCGCAAGACCGGGTCTTT 164  
 QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60  
 DB 165 CTGACCTGGAATGAGTGCAGAGCACAGGTGAGCGGTTTCTCTCTCCAGATTTAGAG 224  
 QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80  
 DB 225 TTTGCCACAGAGATGAGGCTTGGGCTTTGTTCAGGAAATCTGCAAGCGCGGAATTCA 284  
 QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100  
 DB 285 GAGGGCATGAAATCAACATGGAAGATCGAGGCGGCGCAAGACCGAGCTCCGT 344  
 QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyraAlaLysHisMetLysPro 120  
 DB 345 GAGCCACTGGATGAGATGACATGAAGCGCAGAGCGGCTTATGCAAGCACATGAAGCGC 404  
 QY 121 SerValGluProAlaProValSerArgAspThrPheSerTyrrMetGlyAspPheVal 140  
 DB 405 AGCGTGGAGCGCGGCTCCAGTTAGCAGAGACACAGTTTCTCTACATGGGAGACTTCGTC 464

141 ValValTyrThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160  
 Db GTGGTCTACACTGCTGCTCTCCAGTATATGGCGTAGAAGGCGCGAGCAAGATC 524  
 QY 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180  
 Db 525 GGGTTCATCTAGGGGGCGGGCCATCTTTAAATGATAGGCAATTAGACTTCTGGGGCGAG 584  
 QY 181 ThrAsnGlnArgAlaGluIleHisAlaCysLysAlaIleGluGlnAlaLysThrGln 200  
 Db 585 ACAACCAAGAGAGCGGAATTCATGAGCTGCAAGCCATTGAAACAAAGAACTCAA 644  
 QY 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220  
 Db 645 AACATCAATAACTGGTTCGTATACAGACAGTATGTTTACGATAAATGTTATACCTAAC 704  
 QY 221 TrpValGlnGlyTrpLys-LysAsnGlyTrpLysThrSerAla-GlyLysGluValIleA 240  
 Db 705 TGGGTTCAAGTTTGAAGNAATAATGGTGGAGACAAGTGCAGGGGAAAGAGGTGATCA 764  
 QY 240 snLysGluAspPheValAlaLeuGluArgLeuThrGln-GlyMetAspIleGlnTrpMet 259  
 Db 765 ACAAGAGGACTTTGTGGCACTGGAGAGCTTACCCAGGGGATGAGACATTGAGTGGATG 824  
 QY 260 HisValPro-GlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgG 279  
 Db 825 CATGTTCTGGGTCATTCGGATTTATAGGCCATGAAGAACTGACAGATTAGCCAGGNA 884  
 QY 279 uGlyAlaLysGlnSerGluAsp 286  
 Db 885 AGGAGCTAAACCATCGGAAGAC 906

RESULT 4  
 BM810651 1060 bp mRNA linear EST 05-MAR-2002  
 LOCUS AGENCOURT\_6580549 NIH\_MGC\_98 Homo sapiens cDNA clone IMAGE:5453776  
 DEFINITION 5', mRNA sequence.  
 ACCESSION BM810651  
 VERSION BM810651.1 GI:19127474  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1060)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: sgapsb@mail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 Plate: LNCM1947 row: n column: 17  
 High quality sequence stop: 725.  
 Location/Qualifiers  
 1..1060  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5453776"  
 /tissue\_type="astrocytoma grade IV, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH MGC 98"  
 /note="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCACGAG(G). Library constructed by Ling Hong

in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."

ORIGIN  
 Alignment Scores: 2.3e-125 Length: 1060  
 Pred. No.: 1446.00 Matches: 274  
 Score: 1446.00  
 Percent Similarity: 96.86% Conservative: 4  
 Best Local Similarity: 95.47% Mismatches: 8  
 Query Match: 93.53% Indels: 2  
 DB: 12 Gaps: 0

US-10-054-313-1 (1-286) x BM810651 (1-1060)

QY 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20  
 Db 58 ATGAGCTGGCTTCGTTCCTGGCCACAGAGTGGCTTGGCGCTTGGCCCTGCGCGC 117  
 QY 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyValGlyThrGlyValPhe 40  
 Db 118 GCCTCTCGCGGTTTCGGGATGTTCTATGCCGTGAGGAGGGGCCCAAGACCGGGGCTTT 177  
 QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLys 60  
 Db 178 CTGACCTGGAATGAGTGCAGAGCACAGTGGACCGGTTTCTGCTGCCAGATTTAAGAAG 237  
 QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80  
 Db 238 TTTGCCACAGAGGATGAGGCTTGGGCTTGTTCAGGAAATCTGCAAGCCCGGAAGTTTCA 297  
 QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100  
 Db 298 GAAGGCGATGAAATCAACATGGACAAGATCGAGGCGAAGCCAGCAGCGACTCCGT 357  
 QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120  
 Db 358 GAGCCACTGGATGGAGATGGACATGAAAGCGCAGAGCCGATATGCAAGCCAGCAATGAACCG 417  
 QY 121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140  
 Db 418 AGCGTGGAGCGCGGCTCCAGTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC 477  
 QY 141 ValValTyrThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160  
 Db 478 GTCGCTACACTGATGCTGCTGCTCCAGTAATGGCGGTAGAGGCGCGCAGCAGGAATC 537  
 QY 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180  
 Db 538 GCGGTTTACTGGGGCGGGCCATCTTTAAATGTAGGCATTAGACTTCTGGGGCGGAG 597  
 QY 181 ThrAsnGlnArgAlaGluIleHisAlaCysLysAlaIleGluGlnAlaLysThrGln 200  
 Db 598 ACAACCAAGAGAGCGGAATTCATGAGCTGCAAGCCATTGAAACAAAGCAAGAACTCAA 657  
 QY 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220  
 Db 658 AACATCAATAACTGGTTCGTATACAGACAGTATGTTTACGATAAATGTTATACCTAAC 717  
 QY 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240  
 Db 718 TGGGTTCAAGGTTTGAAGAA-ATGGGTGGAAGACAAGTGCAGGGAAGAGGTGATNAC 776  
 QY 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260  
 Db 777 AAAGAGGACTTTGTGGCACTGGAGGCTTACCCAGGGGATGGACATTCAGTGGATGCAT 836  
 QY 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArg-LeuAlaArgGluG 280  
 Db 837 GGTCTGCTCATTCGGGATTTATANGGCAATGAAGAGCTGACAGATTATCCAGAGAAGG 896  
 QY 280 yAlaLysGlnSerGluAsp 286

```

Db      897 AGCTNAACAATCGGAAGAC 915
|||||
AL554334      952 bp      mRNA      linear      EST 31-MAY-2003
AL554334      Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
LOCUS      clone CS0DI082YL10 5-PRIME, mRNA sequence.
DEFINITION
ACCESSION      AL554334
VERSION
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS      Li W.B., Gruber C., Jessee, J. and Polayes, D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL      Unpublished (2001)
COMMENT      On Feb 15, 2001 this sequence version replaced gi:12895013.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 1105.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DI082DF05QPL&cluster=1105.r. Contact :
Feng liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DI082DF05QPL.
LOCATION/Qualifiers
1. 952
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI082YL10"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Alignment Scores:
Pred. No.:      4,046-124      Length:      952
Score:      1432.00      Matches:      263
Percent Similarity: 99.25%      Conservative: 2
Best Local Similarity: 98.50%      Mismatches: 0
Query Match:      92.63%      Indels: 0
DB:      9      Gaps: 0
US-10-054-313-1 (1-286) x AL554334 (1-952)
QY      1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db      152 ATGAGCTGGCTTCTTCTCTGCCCACAGAGTCGCTTGGCGGCTTGCCTGCCCGCCG 211
QY      21 GlySerArgGlyPheGlyMetPheTyzAlaValArgGlyArgGlyThrGlyValPhe 40
Db      212 GGTCTCGCGGTTTCGGGATGTTCTATCCGTGAGAGGGGCCGCAAGACCGGGTCTTT 271
QY      41 LeuThrTrpAnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheValys 60
Db      272 CTGACCTGGAATGAGTCAGAGCACAGGTGACCGGTTCTGCTGTCAGATTTAAGAAG 331
QY      61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db      332 TTGGCCACAGAGGATGAGCCCTGGGCTTTGTCAGAAATCTGCAAGCCCGGAAGTTTCA 391
QY      81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
|||||
392 GAAGGCGCATGAAATCAACATGGACAGAAATCGGAGGCGAAGCCAGCGACTCCGT 451
QY      101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyzAlaLysHisMetLysPro 120
Db      452 GAGCCACTGGATGGAGATGGATGAAGCGGAGAGCCGTATGCAAGCACATGAAGCCG 511
QY      121 SerValGluProAlaProValSerArgAspThrPheSerTyzMetGlyAspPheVal 140
Db      512 AGCGTGGAGCGCGCGCTCCAGTTAGCAGAGACAGCTTTTCTCATCTGGGAGACTTCGTC 571
QY      141 ValValTyzThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160
Db      572 GTGCTGTACACTGATGGCTGCTGCTCCAGTAATGGCGTAGAAGCGCGGAGCAGGAATC 631
QY      161 GlyValTyzTrpGlyProGlyHisProLeuAsnValGlyLeuArgLeuProGlyArgGln 180
Db      632 GCGGTTTACTGGGGCCAGGCCATCTCTTAAATGTAGGCATTAGACTTCTCTGGGCGGCGAG 691
QY      181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200
Db      692 ACAACCAAGAGCGGAAATTCATGCAGCTGCAAGCCATTGAACAAGCAAGAACTCAA 751
QY      201 AsnIleAsnLysLeuValLeuTyzThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db      752 AACATCAATAAACTGGTCTCTATACAGACAGATATGTTTACGATAAATGGTATACTAAC 811
QY      221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysTyzSerAlaGlyLysGluValIleAsn 240
Db      812 TGGGTTCAAGTTTGGAGAAATGGTGGAGACAGATGCGAGGAGAAAGAGGTGATCAAC 871
QY      241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Db      872 AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCAGGGGATGGACATTCAGTGGATGCAT 931
QY      261 ValProGlyHisSerGlyPhe 267
Db      932 GTTCCCTGGTCATCGGGAATTT 952
RESULT 6
BG829086      944 bp      mRNA      linear      EST 22-MAY-2001
LOCUS      602752428F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4905246 5',
DEFINITION      mRNA sequence.
ACCESSION      BG829086
VERSION      BG829086.1 GI:14176673
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI806 row: g column: 07
High quality sequence stop: 781.
Location/Qualifiers
1. 944
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4905246"
/tissue_type="rhabdomyosarcoma"
FEATURES
source

```

```

/lab_host="DH108 (phage-resistant)"
/clone_lib="NH108"
/notes="Organ: muscle; Vector: pOTB7; Site 1: EcoRI;
Site 2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

## ORIGIN

```

Alignment Scores:
Pred. No.: 1,42e-122 Length: 944
Score: 1415.50 Matches: 271
Percent Similarity: 96.18% Conservative: 6
Best Local Similarity: 94.10% Mismatches: 8
Query Match: 91.56% Indels: 4
DB: 12 Gaps: 1
```

US-10-054-313-1 (1-286) x BG829086 (1-944)

```

QY 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
DB 4 ATGAGCTGCTTCTCTTCTGCGCCACACAGTCGCTTGGCGCCCTTGCCTGCGCGCGC 63
QY 21 GlySerArgGlyPheGlyMetPheTyxAlaValArgGlyArgGlyValPhe 40
DB 64 GGTCTTCGCGGTTTGGGATGTTCTATGCCGTGAGAGGGGGCGGACCGGGTCTTT 123
QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLys 60
DB 124 CTGACCTGAATGAGTGACAGACACAGGTGGACCGGTTTCTCTGTCGCCAGATTAAGAAG 183
QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
DB 184 TTTCGCACAGAGATGAGCGCTTGGCGCTTTGTCAGAAATCTGCAAGCCGCGAATTTCA 243
QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
DB 244 GAAGGCGATGAATCAACATGACATGACAGAAATCGGAGGCGAAGCCAGCAGCTCGT 303
QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyxAlaLysHisMetLysPro 120
DB 304 GAGCCACTGATGGAGATGACATGAATGAGAGCGCGTATGCAAGACACATGAAGCGC 363
QY 121 SerValGluProAlaProValSerArgAspThrPheSerTyxMetGlyAspPheVal 140
DB 364 AGCGTGGAGCGCGGCTTCAGTTAGCAGACACAGTTTCTCTATGAGGACTTCGTC 423
QY 141 ValValTyxThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyTle 160
DB 424 GTGCTTACACTGATGGCTGCTCCAGTAATGGCGGTAGAGGCGCGAGCAAGAAATC 483
QY 161 GlyValTyxTrpGlyProGlyHisProLeuAsnValGlyLeuArgLeuProGlyArgGln 180
DB 484 GGGTTTACTGGGGCCAGGCCATCTTTAAATGATAGCATTAGACTTCTCTGGCGGGCAG 543
QY 181 ThrAspGlnArgAlaGluLeuHisAlaLysLysAlaLysLysAlaLysLysGluVal 200
DB 544 ACAAAACCAAGAGCGGAAATTCATGCGGCTGCAAAAGCCATTGAACAGCAAGACTCA 603
QY 200 nAsnIleAsnLysLeuValLeuTyxThrAspSerMetPheThrIleAsnGlyLysThrAs 220
DB 604 AAACATCAATAAATCTGTTCTGATACACAGACAGTATGTTTACGATAATGGTATTAACAA 663
QY 220 nTrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValLys 240
DB 664 CTGGGTTCAAGGTTGGAAGAAAATGGGTGGAGACAAAGTGCAGCAAGAGAGGTGATCAA 723
QY 240 nLysGluAspPheValAla-LeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetH 260
DB 724 CAAAGAGGACTTTGTGGCACTTGGAGAGCTTACCAAGGGGATGGACATTCAGTGGATGC 783
```

```

QY 260 isValProGlyHisSerGlyPheIleGlyAsnGluAlaAspArgLeuAlaArgGluG 280
DB 784 ATGTACCGGTCATCGGATTAT--AGGCATGAAGAAGCTGACAGATTAGCCAGAGAAG 840
QY 280 lyAlaLysGlnSerGluAsp 286
DB 841 GAGTTAA-CAATCGAAGAC 859
RESULT 7
BX416456
LOCUS
DEFINITION
  BX416456 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
  CS0DA007Y012 5-PRIME, mRNA sequence.
ACCESSION
  BX416456
VERSION
  BX416456.1 GI:30763629
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1. (bases 1 to 1196)
  Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
  Full-length cDNA libraries and normalization
  Unpublished (2001)
  Contact: Genoscope
  Genoscope - Centre National de Sequencage
  BP 131 91006 EVRY cedex - France
  Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
  Library was constructed by Life Technologies, a division of
  Invitrogen. This sequence belongs to sequence cluster 1105.r For
  more information about this cluster, see
  http://www.genoscope.cns.fr/
  cgi-bin/cluster.cgi?seq=CS0DA007BH06QPI&cluster=1105.r. Contact :
  Feng Liang Email : fliang@lifetech.com URL :
  http://fulllength.invitrogen.com/ invitrogen Corporation 1600
  Faraday Avenue Genoscope sequence ID : CS0DA007BH06QPI.
  Location/Qualifiers
    1..1196
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="CS0DA007Y012"
      /tissue_type="NEUROBLASTOMA"
      /notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed
      with a NotI-oligo(dT) primer. Five prime end enriched,
      double-strand cDNA was digested with Not I and cloned into
      the Not I and EcoRV sites of the pCMVSPORT 6 vector.
      Library was not normalized."
```

## ORIGIN

```

Alignment Scores:
Pred. No.: 1,61e-121 Length: 1196
Score: 1406.00 Matches: 277
Percent Similarity: 97.55% Conservative: 2
Best Local Similarity: 96.85% Mismatches: 7
Query Match: 90.94% Indels: 4
DB: 13 Gaps: 0
US-10-054-313-1 (1-286) x BX416456 (1-1196)
```

```

QY 1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
DB 125 ATGAGCTGCGTCTTGTTCCTGGCGCCACAGAGTCGCTTGGCGCCCTTGCCTGCCCGCC 184
QY 21 GlySerArgGlyPheGlyMetPheTyxAlaValArgGlyArgLysThrGlyValPhe 40
DB 185 GGCTCTCGCGGTTGCGGATGTTCTATGCCGTGAGAGGGGGCGCAAGACCGGGTCTTT 244
QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLys 60
DB 245 CTGACCTGGAATGAGTCAGAGCACAGGTCGACCGGTTTCTCTGCTCCAGATTTAAGAAG 304
```

QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80  
 Db 305 TTTCACACAGAGATAGGCGCTTGTGTCAGAAATCTGCAAGCCGGAAGTTTCA 364  
 QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100  
 Db 365 GAAGGCGCATGAATCAATCAGATGACAGAAATCGGAGGCGGAAGC-AGCAGCGACTCCGT 423  
 QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120  
 Db 424 GAGCCACTGGATGGAGATGACATGAAGCGCAGACGGTATGCAAGCACATCAAGCCG 483  
 QY 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140  
 Db 484 AGCGTGGAGCGCGCCCTCCAGTTAGCAGAGACACGTTTCTTCACTAGGAGACTTCGTC 543  
 QY 141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160  
 Db 544 GTCGCTACACTGATGGCTGCTGCTCCAGTAATGGCGTAGAGCGCGGAGGATC 603  
 QY 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180  
 Db 604 GCGCTTTACTGGGGCGCCAGGCCATCCTTTAAATGTAGCATTAGACTTCTGGCGGCGAG 663  
 QY 181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200  
 Db 664 AAAAAACCAAGAGCGGAATTCATGCGCTGCAAGCCATTGAACAGCAAGACTCAA 723  
 QY 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220  
 Db 724 AACATCAATAAACTGGTTCTGTATACACAGACAGTAGTGTTCAGATAAATGGTATAACTAAC 783  
 QY 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240  
 Db 784 TGGGTTCAAGTTTGAAGAA-AATGGGTGGAGAACAGAGTGCAGGAAGAGGTGATCAC- 841  
 QY 241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260  
 Db 842 AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAGTGGATGCAT 901  
 QY 261 ValProGlyHisSerGlyPheIleGlyAsnGluAlaLysArgLeuAlaArgGluGly 280  
 Db 902 GTYCT-GGTCACTCGGATTTATAGGCAATGAAGAGCTGACAGATTAGCCAGAGAGGA 960  
 QY 281 AlaLysGlnSerGluAsp 286  
 Db 961 GCTAAACAATCGGAAGAC 978

RESULT 8  
 BUI75613  
 LOCUS  
 DEFINITION BUI75613 893 bp mRNA linear EST 04-SEP-2002  
 AGENCOURT 7935024 NIH\_MGC\_68 Homo sapiens cDNA clone IMAGE:6009671  
 5', mRNA sequence.  
 ACCESSION BUI75613  
 VERSION BUI75613.1 GI:22689597  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 893)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: DCTD/BTP/Gazdar  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov  
 Plate: L1AM13196 row: h column: 24  
 High quality sequence stop: 672.  
 FEATURES  
 Location/Qualifiers  
 source  
 1..893  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6009671"  
 /tissue\_type="large cell carcinoma"  
 /lab\_host="DH109 (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_68"  
 /notes="Organ: lung; Vector: pCMV-SPORT6, Site: 1: NotI,  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.  
 Average insert size 1.8 kb. Library constructed by Life  
 Technologies."

## ORIGIN

Alignment Scores:  
 Pred. No.: 5,61e-116 Length: 893  
 Score: 1345.00 Matches: 270  
 Percent Similarity: 95.14% Conservative: 4  
 Best Local Similarity: 93.75% Mismatches: 4  
 Query Match: 87.00% Indels: 10  
 DB: 13 Gaps: 0  
 US-10-054-313-1 (1-286) x BUI75613 (1-893)  
 QY 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20  
 Db 27 ATGAGCTGGCTTCTGTTCTGCTGCCACAGAGTCGCTTGGCGCCTTGGCTCGCGCCG 86  
 QY 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40  
 Db 87 GCGCTCTCGCGGTTCGGGATGTTCTATGCCGTGGAGGGCGCGAAGCCGGGTCTTT 146  
 QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60  
 Db 147 CTGACCTGGATGAGTGCAGACACAGGTGGACCGGTTTCTGCTGCCAGATTGAAGAG 206  
 QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80  
 Db 207 TTTCACACAGAGATAGGCGCTTGTGTCAGAAATCTGCAAGCCGGAAGTTTCA 266  
 QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100  
 Db 267 GAAGGCGCATGAATCAATCAGATGACAGAAATCGGAGGCGGAAGCCAGCGACTCCGT 326  
 QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120  
 Db 327 GAGCCACTGGATGGAGATGACATGAAGCGCAGACGGTATGCAAGCACATCAAGCCG 386  
 QY 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140  
 Db 387 AGCGTGGAGCGCGCCCTCCAGTTAGCAGAGACACGTTTCTTCACTAGGAGACTTCGTC 446  
 QY 141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160  
 Db 447 GTCGCTACACTGATGGCTGCTGCTCCAGTAATGGCGTAGAGCGCGGAGCGGAATC 506  
 QY 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180  
 Db 507 GCGCTTTACTGGGGCGCCAGGCCATCCTTTAAATGTAGCATTAGACTTCTGGCGGCGAG 566  
 QY 181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200  
 Db 567 AAAAAACCAAGAGCGGAATTCATGCGCTGCAAGCCATTGAACAGCAAGACTCAA 626  
 QY 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220  
 Db 627 AACATCAATAAACTGGTTCTGTATACACAGACAGTAGTGTTCAGATAAATGGTATAACTAAC 686  
 QY 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240



Db 687 TGGGTTCAAGTTGGAGAAAATGGTGGAGACAAAGTGCAGGAAAGAGGTGATCAAC 746

Qy 241 LysGluAaspPhe-ValAlaLeu-GluArgLeuThrGlnGly--MetAspIleGlnTrp-M 259

Db 747 AAAGAGGACTTTGGTGGCACTGGGAGAAAGCTTACCCAGGGGATGGGACATTCAGTGGGA 806

Qy 259 ethis-ValProGly-His-SerGlyPheIle-GlyAsn-GluGluAlaAspArgLeuAl 277

Db 807 TGCATGGTTCTGTCGTCATTCGGGATTTATAGGCAATGGAAGAACTGACGATTAGC 866

Qy 277 aArg 278

Db 867 CCAG 870

RESULT 9

BG831321

LOCUS 602786160F1 NIH\_MGC\_42 Homo sapiens cDNA clone IMAGE:4908326 5',

DEFINITION mRNA sequence.

ACCESSION BG831321

VERSION BG831321.1

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 949)

TITLE NIH-MGC http://mgi.nci.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
DNA sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Plate: LLCM1814 row: 9 column: 15  
High quality sequence stop: 807.  
Location/Qualifiers  
1. 949  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4908326"  
/tissue\_type="epithelioid carcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 42"  
/note="Organ: pancreas; Vector: pOTB7; Site: 1: XhoI; Site: 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC library. |"

ORIGIN

Alignment Scores:

Pred. No.: 1.82e-115 Length: 949

Score: 1340.00 Matches: 257

Percent Similarity: 93.93% Conservat: 6

Best Local Similarity: 91.79% Mismatches: 15

Query Match: 86.68% Indels: 3

DB: 12 Gaps: 0

US-10-054-313-1 (1-286) x BG831321 (1-949)

Qy 1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgarg 20

Db 30 ATGAGCTGGCTTCCTGCTCCGCCACAGAGTCGCCCTTGGCCGCTTGCCTGCCGCCG 89

Qy 21 GlySerArgGlyPheGlyMetPheTrpAlaValArgArgGlyArgGlyThrGlyValPhe 40

Db 90 GGCCTCTCGCGGTTTCGGGATGTTCTATGCCGTGAGAGGGGCCCGACAGCCGGGCTCTT 149

Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheGlyLys 60

Db 150 CTGACCTGGAAATGAGTCAGAGCACAGTGGACCGGTTTCTGCTGCAGATTTAAAGAAG 209

Qy 61 PheAlaThrGluAaspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80

Db 210 TTTGCCACAGAGGATGAGGCTGGGCTTGTACAGGAATCTGCAAGCCCGGAATTTCA 269

Qy 81 GluGlyHisGluAsnGlnHisGlyGlnSerGluAlaLysProGlyLysArgLeuArg 100

Db 270 GAAGCGCATGAAAAATCAACATGGCAAGAAATCGGAGCGCAAGCCAGCAAGCGACTCCGT 329

Qy 101 GluProLeuAaspGlyAaspGlyHisGluSerAlaGlnProTrpAlaLysHisMetLysPro 120

Db 330 GAGCCACTGGATGGAGATGGACATGAAGCGCAGAGCCGTATGCNAAGCACATGAAGCCG 389

Qy 121 SerValGluProAlaProProValSerArgAspThrPheSerTrpMetGlyAaspPheVal 140

Db 390 AGCCTGGAGCGCGCCCTCCAGTTAGCAGAGACACAGTTTCTTACATGGGAGACTTCGTC 449

Qy 141 ValValTrpThrAaspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyLe 160

Db 450 GTGCTCTACACTGATGCTGCTGCTCAGTAATGGGCGTAGAAGCGCCGAGCAGGAAATC 509

Qy 161 GlyValTrpTrpGlyProGlyHisProLeuAsnValGlyLeuArgLeuProGlyArgGln 180

Db 510 GGGCTTTACTGGGGCGCGGCCATCTTTAAATGTAGGCATTAGACTTCCTGGCGCGCAG 569

Qy 181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200

Db 570 ACAACCAAGAGCGGAAATTCATGCAGCCTGCAAGCCATTGAACCAAGCAAGCAAGACTCAA 629

Qy 201 AsnIleAsnLysLeuValLeuTrpThrAaspSerMetPheThrIleAsnGlyIleThrAsn 220

Db 630 AACATCAATAAAGTCTGTTCTGTATACAGACAGATGTTTACGATAAATGTTATAACTAAC 689

Qy 221 TrpValGlnGlyTrp-LysLysAsn-GlyTrpLysThrSerAlaGlyLysGluValIleA 240

Db 690 TGGGTTCAAGGTTGGGAGAAAAATGGGTTGGAGAACAAAGTCGAGGCAAGAGGTGATC 749

Qy 240 snLysGluAaspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMeth 260

Db 750 AAAAAGAGGACTTTGTGGGACTGGAGAGGTT-ACCAGGGGATGGGACATTCAGTGGATGC 808

Qy 260 isValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAaspArgLeuAlaArg 278

Db 809 ATGTCCTGGGACATCCGGGAATTATAGGAATGAAGGAGTGAAGGATTAGCGAGA 864

RESULT 10

CA775004

LOCUS ipig06.y1 HR85 islet Homo sapiens cDNA clone IMAGE:6216755 5',

DEFINITION similar to TR:060523 060523 RIBONUCLEASE H TYPE II. [2] TR:060857 ;, mRNA sequence.

ACCESSION CA775004

VERSION CA775004.1

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 765)  
Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K., Lemishka, I., Searce, M., Breetelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wyllie, T., Martin, J., Blaisstein, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J.,

Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R.,  
Williams,T., Jackson,Y. and Bowers,Y.  
Endocrine Pancreas Consortium  
Unpublished (2000)  
Other ESTs: ip11g05.xl  
Contact: Douglas Meiton, Klaus H. Kaestner, & Hiroshi Inoue  
Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmeiton@biohph.harvard.edu

Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:  
Washington University Genome Sequencing Center For information on  
obtaining a clone please contact: Dr. Hiroshi Inoue  
(hinoue@im.wustl.edu)  
Seg primer: -40Rp from Gibco  
High quality sequence stop: 469.  
Location/Qualifiers

1. .765  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6216755"  
/tissue\_type="Purified pancreatic islet"  
/lab\_host="DH10B"  
/clone\_lib="HR85 islet"  
/notes="Organ: Pancreas; Vector: pBluescript SK(-); Site 1:  
NotI; Site 2: XhoI; cDNA made by oligo-dr priming.  
Size-selected on agarose gel. Average insert size -1kb. 5'  
XhoI site was destroyed after directional cloning.  
Amplified once. Contact information: Hiroshi Inoue, MD,  
Metabolism Div. (Alan Permutt Lab), Washington University  
School of Medicine, Box 8127, 660 South Euclid Ave., St.  
Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:  
314-362-1916, Fax: 314-747-2692."

ORIGIN

Alignment Scores:  
Pred. No.: 1.02e-110 Length: 765  
Score: 1288.00 Matches: 237  
Percent Similarity: 99.17% Conservatve: 3  
Best Local Similarity: 97.93% Mismatches: 2  
Query Match: 83.31% Indels: 0  
DB: 14 Gaps: 0

US-10-054-313-1 (1-286) x CM775004 (1-765)

Qy 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20  
Db 40 ATGAGCTGGCTTCTGTTCTGGCCACAGAGTCGCTTGGCCCTTGGCCCTGCCCGCCG 99  
Qy 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40  
Db 100 GCGTCTCGGGTTCGGGATGTTCTATGCGGTGAGGAGGGCCGAGACCGGGGTCTTT 159  
Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60  
Db 160 CTGACCTGGAATGAGTCAGACAGACAGTCGACCGGTTTCTCTGCTCCAGATTATAGAG 219  
Qy 61 PheAlaThrGluaspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80  
Db 220 TTGCGCACAGAGATGAGGCTGGGCTTTGTGAGGAAATCTCAAGCCCGGAGTTTCA 279  
Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLysArg 100  
Db 280 GAAGGCGATGAAATCAACATGACATGACATGACATGACATGACATGACATGACATG 339  
Qy 101 GluProLeuaspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120  
Db 340 GAGCCACTGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATG 399

Qy 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140  
Db 400 ACGGTGGAGCGCGCCCTCCAGTTAGCAGACAGACAGTTTCTCTACATGGAGACTTCGTC 459  
Qy 141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyLe 160  
Db 460 GTCTGCTACACTGATGGTGTCTCCAGTATATGGCGGTAGAGGCGCGAGCAGGAATC 519  
Qy 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyLleArgLeuProGlyArgGln 180  
Db 520 GCGGTTTCTGGGGCCAGGCCATCTTTAATGATGAGCATTTAGACTTCTCTGGGGCGGAG 579  
Qy 181 ThrAsnGlnArgAlaGluGluHisAlaAlaCysLysAlaLleGluGlnAlaLysThrGln 200  
Db 580 ACAACCAAGAGCGGAAATTCATGAGCTGCAAGCAATGAAACAGCAAGCAAGCAACTCAA 639  
Qy 201 AsnLleAsnLysLeuValLeuTyrThrAspSerMetPheThrLleAsnGlyLleThrAsn 220  
Db 640 AACATCAATAAACTGGTCTGTATACAGACAGTATGTTTACGATAAATGGGTATAACTAAC 699  
Qy 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValLleAsn 240  
Db 700 TGGGTTCAAGGTTGGAAGAAATGGGTGGAAGACACAGTGCAGGGAAGAGGTGATCCAC 759  
Qy 241 LysGlu 242  
Db 760 AAAGAG 765

RESULT 11  
BQ921568  
LOCUS  
DEFINITION  
BQ921568  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BQ921568 935 bp mRNA linear EST 20-AUG-2002  
AGENCOURT\_8922346 NIH\_MGC\_71 Homo sapiens cDNA clone IMAGE:6470333  
5', mRNA sequence.  
BQ921568  
EST  
BQ921568.1 GI:22336599  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILNL at:  
http://image.llnl.gov  
Plate: LLAM14000 row: 0 column: 06  
High quality sequence stop: 563.  
Location/Qualifiers  
1. .935  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6470333"  
/tissue\_type="leiomyosarcoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 71"  
/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2.1 kb."

FEATURES  
source

ORIGIN

Alignment Scores:  
Pred. No.: 6.33e-110 Length: 935  
Score: 1281.00 Matches: 255  
Percent Similarity: 93.17% Conservatve: 4

Best Local Similarity: 91.73% Mismatches: 11  
Query Match: 82.86% Indels: 8  
DB: 13 Gaps: 0  
US-10-054-313-1 (1-286) x BQ921568 (1-935)

QY 1 MetSerTrpLeuPheLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20  
DB 98 ATGAGCTGGCTTCTCTCTGCGCCACAGAGTCGCTTGGCGCCCTTCCCTTGGCGCGC 157  
QY 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyArgGlyThrGlyValPhe 40  
DB 158 GGCCTTCGCGGTTTGGGATGTTCTATGCCGFGAGAGGGCGCCGAGACCGGGGTCTTT 217  
QY 41 LeuThrTrpAsnGlyCysArgAlaGlnValAspArgPheProAlaAlaArgPheLys 60  
DB 218 CTGACCTGGAATGAGTCAGAGACAGAGTCGACCGGTTCTCTGCTGCCAGATTTAAGAAG 277  
QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80  
DB 278 TTGTCACAGAGATGAGCGCTGGGCCCTTTGTCAGGAATCTGCAAGCCCGGAATTTCA 337  
QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100  
DB 338 GAAGGCGATGAAATCAACATGACAGATCGAGAGCGAAGCCAGCAAGCGATCGT 397  
QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120  
DB 398 GAGCACTGGATGGAGATGGATGAATGAAGCGCAGAGCGGTATGCAAGACACATGAAGCG 457  
QY 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140  
DB 458 AGCGTGGAGCGCGCTCCAGTTAGCAGAGACACGTTTCTACATGGAGACTTCGTC 517  
QY 141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyLe 160  
DB 518 GTGCTCTACATGATGGCTGCTGCTCCAGTAATGGCGGTAGAGCCCGCAGCAGGAATC 577  
QY 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyLysArgLeuProGlyArgGln 180  
DB 578 GCGGTTACTGGGGCGCGGCCATCTTTANATGTAGGACTTAGACTTCTCTGGCGGCGAG 637  
QY 181 ThrAsnGluArgAlaGluLeuHisAlaLysLysAlaLysLysAlaLysGluAlaLysThrGln 200  
DB 638 ACAACCAAGAGCGGAATTCATGCGCTGCAAGCCATTAACCAAGCAAGCAAGCAACTCAA 697  
QY 201 AsnIleAsn-LysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyLysThrAs 220  
DB 698 ACATCATTAATACTGTTCTGTATACAGACAGTATGTTACATTAATGGTATTAACATAA 757  
QY 220 nTrpValGlnGlyTrpLys-LysAsnGlyTrpLysThrSerAla--GlyLysGluValI 239  
DB 758 CTGGGTTTCAGGTTGGAAGAAATAATGGGTGGGAAGACAGTGCANGGGGAAAAGAGTGT 817  
QY 239 eAsnLysGluAspPheValAlaLeu-GluArgLeuThrGlnGlyMet-AspIleGlnTrp 258  
DB 818 CAACAAGAGGACTTTGTGGCACTGGGAGAGCTTAACCCAGGGAATGGACATCAGNNGN 877  
QY 259 MetHis-ValProGlyHis-SerGlyPheIleGlyAsn 270  
DB 878 ATGATGATGTTCTGGGCAATTCGGGATTAATAGGCCAT 915

RESULT 12  
BE746490  
LOCUS  
DEFINITION  
601579445F1 NIH\_MGC\_9 Homo sapiens cDNA clone IMAGE:3928402 5',  
mRNA sequence.  
ACCESSION  
BE746490  
VERSION  
BE746490.1 GI:10160482  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 779)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs@remail.nih.gov  
Tissue Procurement: DCTD/DTF  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: LLCW759 row: 1 column: 11  
High quality sequence stop: 746.  
Location/Qualifiers  
1..779  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3928402"  
/tissue\_type="adenocarcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/notes="Oligo: ova; Vector: pOTB7; Site 1: XhoI; Site 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGACAGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN  
Alignment Scores:  
Pred. No.: 3,83e-106 Length: 779  
Score: 1239.50 Matches: 238  
Percent Similarity: 95.63% Conservative: 8  
Best Local Similarity: 94.44% Mismatches: 8  
Query Match: 80.17% Indels: 4  
DB: 10 Gaps: 1

US-10-054-313-1 (1-286) x BE746490 (1-779)  
QY 18 CysArgArgGlySerArgGlyPheGlyMetPheTyrAlaValArgGlyArgLysThr 37  
DB 2 TGGCGCGG--TCTCGCGGTTTCGGATGTTCTATGCGGTGAGGAGGCGCGCAAGACC 58  
QY 38 GlyValPheLeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArg 57  
DB 59 GGGGTCTTTCTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTCTCTGCTGCCAGA 118  
QY 58 PheLysLysPheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerPro 77  
DB 119 TTTAAGAAGTTTGCACAGAGGATGAGGCTGGCCCTTTGTCTAGGAAATCTGCAAGCCCG 178  
QY 78 GluValSerGluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLys 97  
DB 179 GAAGTTTTCAGAGGCGCATGAAATCAACATGGAATCGGAGCGGCAAGCCAGCAAG 238  
QY 98 ArgLeuArgGluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHis 117  
DB 239 CGACTCGCTGAGCCACTGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGG 298  
QY 118 MetLysProSerValGluProAlaProProValSerArgAspThrPheSerTyrMetGly 137  
DB 299 ATGAAGCGGAGCGTGGAGCGGCGCTCCAGTTAGCAGAGACACGTTTCTCATATGGGA 358  
QY 138 AspPheValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArg 157  
DB 359 GACTTCGCTGCTGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 418  
QY 158 AlaGlyIleGlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuPro 177

```

Db 419 GCAGGATCGCGTTACTGGGGCCAGGCCATCTTTAATGTAGGCATTAGACTTCCT 478
Qy 178 GlyArgGlnThrAsnGlnArgAlaGluHisAlaLysLysAlaLysAlaLysGluAla 197
Db 479 GGGGGGACAGAAACCAAGAGCGAAATTCATGAGCCTGCAAGGCATTGAACAAGCA 538
Qy 198 LysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGly 217
Db 539 AAGACTCAAAACATCAATAACTGGTCTGTATACAGACATGATGTATTACGATAAATGGT 598
Qy 218 IleThrAsn-TrpValGlnGlyTyrLysLysAsnGlyTyrLysThrSerAlaGlyLysG 237
Db 599 ATAACAACTGGGTTCAGGTTGGAAGAAAATGGGTGGAAGCAAGATGCGAGGAAAGA 658
Qy 237 uValIleAsnLysGluAspPheValAlaLeuGlu-ArgLeuThrGlnGlyMetAspIleG 257
Db 659 GGTGATCAAAAGAGGACTTGTGGCACTGGAGAGGCTTACCAGGGGATGACATTC 718
Qy 257 InTPMetHisValProGlyHisSerGlyPhe 267
Db 719 AGTGGATGCT-GTTCCTGTCATTCGGGATTT 749

RESULT 13
BUS95053 818 bp mRNA linear EST 20-SEP-2002
LOCUS AGENCOURT_8844002 NIH_MGC_142 Homo sapiens cDNA clone IMAGE:6450776
5', mRNA sequence.
ACCESSION BUS95053
VERSION BUS95053.1 GI:23246812
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 818)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: NCI
cDNA Library Preparation: Michael Brownstein Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCW2621 row: p column: 09
High quality sequence stop: 533.
Location/Qualifiers
1..818
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6450776"
/tissue_type="mixed (pool of 40 RNAs)"
/lab_host="DH10B (T1-phage-resistant)"
/clone_lib="NIH_MGC_142"
/note="Vector: pDMR-LIB; Site 1: SfiI (ggccattatggcc);
Site 2: SfiI (ggccgctctggcc); Double-stranded cDNA was
prepared from a pool of 40 cell line polyA+ RNAs (bladder
- 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -
4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,
kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%,
ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary
gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were
used in cloning as follows:
5'-AAGCAGTGTATCAAGCGACGAGTGGCCATTAGCGCGG-3' and
5'-ATTCTAGAGCGGCGGCGGCACATG-DT(30)NN-3'. Full-length
enriched library was constructed using the Clontech
Creator SMART kit and size-selected to contain the >0.5 kb
size fraction (other fractions present in NIH_MGC_141).

```

```

ORIGIN
Alignment Scores: 4.02e-105 Length: 818
Pred. No.: 1229.00 Matches: 236
Percent Similarity: 96.37% Conservative: 3
Best Local Similarity: 95.16% Mismatches: 5
Query Match: 79.50% Indels: 4
DB: 13 Gaps: 0

US-10-054-313-1 (1-286) x BUS95053 (1-818)
Qy 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
Db 67 ATGAGCTGGCTTCTGTTCTGCGCCACAGAGTCGCTTGGCCGCTTGCCTCGCCGCGC 126
Qy 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyArgGlyThrGlyValPhe 40
Db 127 GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGAGGGCGGCAAGACCGGGGCTTT 186
Qy 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLys 60
Db 187 CTGACCTGGAAATGAGTGCAGACACAGGTGGACCGGTTTCTGCTGCCAGATTAAAGA 246
Qy 61 PheAlaThrGluAspGluAlaTyrAlaPheValArgLysSerAlaSerProGluValSer 80
Db 247 TTTGCCACAGAGATGAGGCTGGGCTTGTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA 306
Qy 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db 307 GAAGGGCATGAATCAATCATGTGACAAAGATCGAGGCCAAAGCCAGCAGGACTCCGT 366
Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db 367 GAGCCACTGGATGGAGATGGACATGAAGAGCGAGACCGCTATGCAAAAGCACATGAAG 426
Qy 121 SerValGluProIleProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
Db 427 AGCGTGGAGCGCGGCTCCAGTTACAGACACACGTTTCTTACATGGAGACTTCGTC 486
Qy 141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160
Db 487 GTGCTCTACACTGATGGCTGCTCCAGTAATGGCGTAGAAGCGCGAGCAGGAATC 546
Qy 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
Db 547 GCGGTTTACTGGGGCGCGGCCATCCTTTAAATGTAGGCATTAGACTTCTCGGCGGCG 606
Qy 181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200
Db 607 ACAACCCAAAGAGCGGAAATTCATGCGGCTGCAGAGCCATTGAACACAGCAAGACTCA 666
Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db 667 AACATCAATAACTGGGTCTGTATACAGACAGTATGTTTACGATAATGTAATACTAAC 726
Qy 221 TrpVal-GlnGlyTyrLysLysAsnGlyTyr-LysThrSerAla-GlyLysGluValIle 239
Db 727 TGGGTTCCAGGTTGGAGNAAATGGGTGGNAGACNAGTGCAGGGGAAGAGGTGATC 786
Qy 240 Asn-LysGluAspPhe 244
Db 787 CACAAAAGAGGACTTT 802

RESULT 14
BUS951915 916 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOURT_6397386 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5492781
5', mRNA sequence.
ACCESSION BM451915
VERSION BM451915.1 GI:18500955

```

Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH\_MGC Library."

```

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 916)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/;
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12115 row: 5 column: 22
High quality sequence stop: 626.
FEATURES
Location/Qualifiers
1..916
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5492781"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 67"
/notes="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Library constructed by Life
Technologies."
ORIGIN
Alignment Scores:
Pred. No.: 2,186-104 Length: 916
Score: 1222.00 Matches: 247
Percent Similarity: 89.08% Conservative: 6
Best Local Similarity: 86.97% Mismatches: 16
Query Match: 79.04% Indels: 15
DB: 12 Gaps: 4
US-10-054-313-1 (1-286) x BM451915 (1-916)
QY 1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg 20
DB 75 ATGAGCTGGCTTCTGTCTGCGCCACACAGAGTCGCTTGGCGCCCTTGCCCTGCCCGCCG 134
QY 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgGlyThrGlyValPhe 40
DB 135 GGCTCTCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGCGCCAGACCGGGGTCTTT 194
QY 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
DB 195 CTGACCTGGAATGAGTCGAGACACAGAGTCGAGCGGTCCTGCTGCCAGATTTAAGAG 254
QY 61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
DB 255 TTGCGCACAGAGATAGGCTCGGCTTGTTCAGGAATCTGCAAGCCGCGAAGTTTCA 314
QY 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
DB 315 GAAGGGGATGAAAATCAACATGACAAAGATCGAGCGGCGAAGCCAGCAAGCGACTCCGT 374
QY 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
DB 375 GAGCCACTGATGAGATGACATGATGAAGCGGACCGGTATGCAAGCACATGAAGCCG 434
QY 121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
DB 435 AGCGTGGAGCGCGCGCTCCAGTTAGCAGAGACACGCTTTTCTTACATGGAGACTTCGTC 494

```

```

QY 141 ValValTyrThrAspGlyCysCysSerSerAsnGlyValArgLysProArgAlaGlyLeu 160
DB 495 GTGCTCTACACTGATGGCTGCTCTCAGTAATGGCGTAGAAGCGCGGAGGAAATC 554
QY 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyLeuArgLeuProGlyArgGln 180
DB 555 GGGGTTTACTGGGGCGCAGCCCATCTTTAAATGTAGGCATTAGACTTCTCGCGCGCAG 614
QY 181 ThrAsnGluArgAlaGluLeuHisAlaAlaCysLysAlaLeuGluGlnAlaLysThrGln 200
DB 615 ACAACCAAGACGCGGAATTCATGAGCTGCAAGCCATTGAACAGCAAGACTCAN 674
QY 201 AsnIleAsn-LysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyLeuThrAs 220
DB 675 AACATCCATAAAACTGGTCTCTGTATACAGACATATGTTTACGATAAATGGTATACTAA 734
QY 220 nTTP-ValGlnGly-TrpLysLysAsn-GlyTTP-LysThrSerAla-GlyLysGluValI 239
DB 735 CTGGGGTTCAGGTTTGGAGAAAAATGGGTGGAGACACGATGCGAGGAAAGAGGTGA 794
QY 239 leAsn---LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnT 258
DB 795 TTCACCAAGAGAGACTTTTGTGGCACTGGAAGGCTCTACCCAGGG-----GGAT 845
QY 258 rpMethis-----ValProGlyHisSerGly---PheIleGlyAsnGluG 272
DB 846 GGACCTTTCCTCCGGGATGCTTGTTCCTCCGGGCACTCCGGGCACTTTTAGGCCAATGAAA 905
QY 272 lu 272
DB 906 AA 907
RESULT 15
LOCUS BF984400
DEFINITION 602307923F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4399301 5',
mRNA sequence.
ACCESSION BF984400
VERSION BF984400.1 GI:12387212
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 869)
NIH-MGC http://mgi.nci.nih.gov/;
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10102 row: n column: 06
High quality sequence stop: 691.
FEATURES
Location/Qualifiers
1..869
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4399301"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_88"
/notes="Organ: small intestine; Vector: pCMV-SPORT6;
Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
oligo-dT primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."

```

## ORIGIN

Alignment Scores: 9.16e-104 Length: 869  
Pred. No.: 1215.00 Matches: 231  
Score: 1215.00  
Percent Similarity: 97.90% Conservative: 2  
Best Local Similarity: 97.06% Mismatches: 4  
Query Match: 78.59% Indels: 2  
DB: 10 Gaps: 0

US-10-054-313-1 (1-286) x BF984400 (1-869)

QY 47 ArgAlaGlnValAspArgPheProAlaAlaArgPheLysLysPheAlaThrGluAspGlu 66  
Db 3 AGAGCACAGGTGGCCGCTTCTCGTCCAGATTAAAGAGTTTGCCACAGAGGATGAG 62  
QY 67 AlaTrpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHisGluAsnGln 86  
Db 63 GCCTGGGCGCTTTGTGAGAAATCTGCAAGCCCGGAAGTTTCAGAAGGGCATGAAAAATCAA 122  
QY 87 HisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGluProLeuAspGlyAsp 106  
Db 123 CATGGACAAGAAATCGAGGGCGAAGCCAGCAAGCGACTCCGTGAGCCACTGGATGGAGAT 182  
QY 107 GlyHisGluSerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaPro 126  
Db 183 GGACATGAAGCGCAGAGCCGCTATGCAAAAGCACATGAAGCCGCGGTGGAGCCGCGCCT 242  
QY 127 ProValSerArgAspThrPheSerTyrMetGlyAspPheValValValTyrThrAspGly 146  
Db 243 CCAGTTACAGAGACACGTTTCTTACATGGAGACTTCGTCTCGTCTACACTGATGGC 302  
QY 147 CysCysSerSerAsnGlyArgArgLysProArgAlaGlyIleGlyValTyrTyrGlyPro 166  
Db 303 TGCTGCTCCAGTAATGGCGGTAGAAGCGCGGAGCAGGAATCGGCGTTTACTGGGGGCGG 362  
QY 167 GlyHisProLeuAsnValGlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGlu 186  
Db 363 GSCCATCCTTTTAAATGTAGGCATTTAGACTTCTGGGGCGGACACAAACCAAGAGCGGAA 422  
QY 187 IleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuVal 206  
Db 423 ATTCATGCAGCGCTGCAGACCATTCACAGCAAGACTCAAAACATCAATAAACTGGTT 482  
QY 207 LeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTyrValGlnGlyTyrLys 226  
Db 483 CTGTATACAGACAGATATGTTTACGATAAATGGTAACTAACTAAGTTCAAGGTGGAAG 542  
QY 227 LysAsnGlyTyrLysThrSerAlaGlyLysGluValIleAsnLysGluAspPheValAla 246  
Db 543 AAAAATGGGTGGAGACACAGTGCAGGAAAGAGGGTGATCAACAAGAGACTTTGTGGCA 602  
QY 247 LeuGluArgLeuThrGlnGlyMetAspIleGlnTyrMetHisValProGlyHisSerGly 266  
Db 603 CTGGAGAGGCTTACCACA-GGGATGGACATTCAGTGGATGCATGTTCTCTGGTCATTCCGGGA 661  
QY 267 PheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyAlaLysGln 283  
Db 662 TTTATAGGCAATGAAGAGCTGACAGATTAGCCAGAGAGGCGCTAAACAA 713

Search completed: March 23, 2004, 00:10:25  
Job time : 2344 secs



Creation date: 06-18-2004  
Indexing Officer: SMOHAMMED - SUAD MOHAMMED  
Team: 1600PrintWorkingFolder  
Dossier: 10469189

Legal Date: 03-12-2004

No.	Doccode	Number of pages
1	WFEE	1

Total number of pages: 1

Remarks:

Order of re-scan issued on .....